

Modern inversion workflow of the multimodal surface wave dispersion curves: Staging strategy and Pattern search with embedded Kuhn-Munkres algorithm

Yingwei Yan,¹ Xiaofei Chen,^{2,1*} Nan Huai³ and Jianbo Guan⁴

¹ Department of Earth and Space Sciences, Southern University of Science and Technology, Shenzhen, 518055, China.

² Shenzhen Key Laboratory of Deep Offshore Oil and Gas Exploration Technology, Southern University of Science and Technology, Shenzhen, 518055, China. Email: chenxf@sustech.edu.cn

³ School of Geomatics and Prospecting Engineering, Jilin Jianzhu University, Changchun, 130118, China.

⁴ College of Geological Engineering and Surveying, Chang'An University, Xian, 710064, China.

SUMMARY

The dispersive information of the higher surface wave modes is beneficial for improving the resolution of the inverted S-wave velocity structure, increasing the penetration depth and enhancing the robustness of the inversion. The recently proposed frequency-Bessel (F-J) transformation can achieve the relatively stable measurement of the higher-mode surface waves. However, sometimes the phenomena of the mode losses and aliases would appear in the dispersion energy image of the seismic records, which may lead to the mode misidentification. And the wrong mode identification is likely to cause the negative impacts on the dispersion inversion and geological interpretation about the survey site. In view of this, we recommend a staging strategy for the inversion of multimodal surface wave dispersion curves. The pattern search (PS) is used to invert the reliable segment of the fundamental-mode surface wave phase velocities for the first stage. For the second stage, the inverted result of the first stage is set as the initial model, the PS with embedded Kuhn-Munkres (PSEKM) algorithm is adopted for inverting the observed phase velocities of all modes. And for each frequency, a weighted bipartite graph is established between the observed values with no-explicitly-specified-mode-order (NESMO) and predicted values of the model \mathbf{m} during the inversion, then the maximum match is determined by the Kuhn-Munkres algorithm for calculating the minimum distance between the observed and predicted data sets. The mode-order information of the observed phase velocities with NESMO would be dynamically evaluated for each model \mathbf{m} occurred in the inversion process. The synthetic reconstruction tests have confirmed the effectiveness of the novel workflow. Also, the

performance of the preconditioned steepest-descent (PSD) algorithm of local optimization methods and influence of the mode misidentification on the inversion result are also clarified in the synthesis tests. The comparison results show that the proposed workflow can realize the nice data fitting and model reconstruction without the time-consuming manual mode-identification for the higher-mode surface waves. Then, the new workflow is applied to the analysis of the actual surface wave data sets collected on two roadbeds, which is still satisfactory. Finally, we discuss the role of the staging strategy of the workflow.

Keywords: Surface wave and free oscillations; Inverse theory; Seismic tomography

1. INTRODUCTION

S-wave velocity is an important parameter to describe the seismic attributes of the shallow subsurface, and it is also related to the dynamic characteristics of the medium (Yilmaz et al. 2006). We can also use the empirical formulas to estimate other properties from the S-wave velocity information of the survey point, such as the porosity (Uyanik, 2019). Moreover, the S-wave velocity is also useful for exploration seismology, such as designing filters to remove the ground roll from seismic reflection records, and the static correction in migration (Mari 1984; Papadopoulou et al. 2020). The surface wave (Rayleigh wave) field observed on the free surface has the strongest sensitivity to the S-wave velocity (Xia et al. 2003; Pan et al. 2019). Nowadays, utilizing the dispersive properties of surface waves to reveal the S-wave velocity structure of the formation (Xia et al. 1999, 2003; Socco et al. 2010; Li et al. 2017) has become the most common method. According to the configuration of seismic sensors, there are two-station-based and multi-station-based dispersion measurement (Hadiouche et al. 1986; Park et al. 1999; Xia et al. 1999). However, there is a 2π ambiguity for the two-station-based method (Pan et al. 2015; Yan et al. 2022). Also, only the fundamental-mode surface wave dispersive curve is extracted, the higher-mode surface waves cannot be perceived (Zywicki & Rix 2005; Maraschini et al. 2010a). The multi-station-based dispersion measurement can not only give the surface wave phase velocities of fundamental-mode but also that of higher-mode, and the dispersive information about higher-mode surface waves is beneficial for improving the resolution of the inverted result (Xia et al. 2003) and reducing the degree of multiple solutions of the inversion (Ernst 2008).

Surface wave propagation is a multimodal phenomenon (Aki & Richards 1980). In other words, the surface wave would travel with several different velocities at each frequency (Aki &

Richards 1980; Xia et al. 2003; Maraschini & Foti 2010b). This phenomenon is manifested by the several peaks at each frequency of the dispersive spectrum, and the spectrum can be provided by different wavefield transformation methods, such as the τ - p transform (McMechan & Yedlin 1981), 2D Fourier transform (Gabriels et al. 1987), phase-shift method (Park et al. 1998), high-resolution linear Radon transform (HRLRT) (Luo et al. 2008), and frequency-Bessel (F-J) transformation (Forbriger 2003; Wang et al. 2019; Li et al. 2021; Xi et al. 2021; Zhou & Chen 2021). Except for the F-J transformation, the other methods are all based on the assumption of the plane-wave propagation. However, the surface wave excited by a point source on the free surface is diffused by the cylindrical wave in real world, which may introduce the model incompatibility errors (the undesirable situation would be verified by the field data example in the following paragraphs) in the measured values defined by the plane-wave-based methods. Forbriger (2003) first analyzed the dispersion of shallow-seismic wavefields by a discrete approximation of Bessel transformation and pointed out that the transform is a complete representation of the data. Then, Wang et al. (2019) presented the frequency-Bessel (F-J) transformation for extracting the surface wave overtones from the ambient noise records, and the transform is essentially derived by the cylindrical wave analysis. The F-J transform has excellent performance in retrieving the higher-mode surface waves, which has been confirmed by the applications with various-scale (Li et al. 2020; Wu et al. 2020). Later, some modifications for the F-J transform were completed by Li et al. (2021), Xi et al. (2021), Zhou and Chen (2021).

It has been widely confirmed that higher-mode data of the surface waves without the wrong mode identification can stabilize the inversion (Maraschini & Foti 2010b), increase the penetration depth and resolution (Xia et al. 2003; Ernst 2008; Pan et al. 2019) and provide the additional information about the interlayers (Pan et al. 2019). However, not all modes of surface waves can get the considerable development in the seismic records, and the energy distribution about the surface wave modes is extremely unbalanced especially when the strata have interlayers (Maraschini et al. 2010a; Maraschini & Foti 2010b). At this time, the mode losses and jump would occur on the f (frequency)- v (velocity) image, which would be easy to lead to the mode misidentification. This is a major pitfall of the dispersion analysis, some researchers have shown that the mode misidentification, especially at low frequencies may bring more than 50% error to the reconstructed structure (O’Neil & Matsuoka 2005). For solving the problem, Ryden and Park (2006) suggested to invert the dispersive spectrum to avoid the explicit judgement of

mode-orders. The extra displacement calculation is required for this method, consequently, it's more computationally expensive. Meanwhile, Lu et al. (2007) also pointed out that when the surface displacement distribution and dispersion characteristics of each mode are considered simultaneously during the inversion, the mode misidentification can be avoided. Certainly, this idea has the disadvantage of the higher computational cost. Maraschini et al. (2010a) adopted a new misfit function based on the Haskell-Thomson matrix method, which allows higher modes to be taken into account without need to associate experimental data point to a specific mode. Maraschini et al. (2010a) tested the new misfit function with steepest-descent and quasi-Newton algorithms, and they indicated that there would be some local minima regions on the hypersurface of the new misfit. So, the initial model of the inversion must be given carefully. To avoid the influence of local minima, Maraschini & Foti (2010b) proposed to use the Monte Carlo method to optimize the new misfit. However, millions of enumerations are necessary for the algorithm, and as the reconstructed parameters increase, the computation-scale would be increased dramatically. Soon afterwards, Cai et al. (2017) adopted the particle swarm optimization to minimize the new misfit, and they also indicated that the wrong mode identification would produce a negative impact to the conventional inversion of the multimodal dispersion curves. Gao et al. (2014) considered to remove the seismic events with higher apparent velocity from the near-source traces to prevent the mode misidentification caused by the leaky wave in low frequency-band of the dispersive image, however, this is a special way without the generalization. Nevertheless, the mode misidentification in high frequency-band is not considered in the approach, and it is still necessary to expand, develop and supplement methods to response the problem.

We know that it's not difficult to pick and identify the most reliable segment of the fundamental-mode surface wave dispersion curve from the dispersive image. Also, the unambiguous identification of the fundamental-mode surface wave is essential condition for the reconstruction of the underground media (Zhang & Chan 2003). So, a seasoned expert might divide the inversion into two-stage when faced with the potential mode misidentification, first, only inverting the reliable fundamental-mode surface wave and calculating the multimodal dispersion curves of the inverted result, then explicitly specifying the mode-orders of the other observed phase velocities (except the measured values of the fundamental-mode) based on the proximity between the predicted multimodal dispersion curves and the dispersive energy

associated with the other observed values. Once the mode-order information of all observed phase velocities has been identified, and the inversion for the second stage is started naturally with the inverted result of the first stage as the initial model. The idea is reasonable and effective in most cases, and it has also been applied in practice (Yang et al. 2019; Li et al. 2022). However, it's time-consuming to manually judge the order of other modes (except the fundamental-mode) for surface waves at the second stage, which may also be inconvenient and error-prone, furthermore, each inversion case for the multimodal dispersion curves would include such an interrupted process. When the number of the survey point of surface waves is huge, this rough manual judgement method must be inefficient. This leads to such a problem, how to use only the information in the data-domain to automatically identify the mode-order of the observed values of the other surface wave modes? In other words, how to evaluate the distance/difference between the observed data (when the measured data set contains the measured value with no-explicitly-specified-mode-order (NESMO)) and modelled data sets?

On the basis of the minimization principle, we always wish the predicted multimodal phase velocities of model \mathbf{m} that appears in the inversion process to have the least distance from the observed ones (some measured values with NESMO). This goal can be achieved through minimizing the “distance” between the predicted and observed values at each frequency. In fact, it's a many-to-many matching problem, which can be solved by the Kuhn-Munkres (KM) (Kuhn 1955; Munkres 1957) algorithm. So, the mode-order information of the observed phase velocities with NESMO at each frequency is dynamically evaluated for each model \mathbf{m} occurred in the inversion process.

Thus, we adopt a staging strategy for the inversion of multimodal surface wave dispersion curves. The entire inversion process is divided into two stages. First stage: Using the pattern search (PS) to invert the reliable segment of the fundamental-mode surface wave dispersion curve to get the 1D S-wave velocity (Song et al. 2009); Second stage: The inverted result of the first stage is set as the initial model. All the observed phase velocities are added to the inversion system, and the PS with embedded Kuhn–Munkres (PSEKM) algorithm is employed to minimize the misfit. A weighted bipartite graph is built between the observed values with NESMO and predicted values (if the measured value of the fundamental-mode surface wave is existent at the frequency point, the predicted values used here refer to the predicted values except for the fundamental-mode, otherwise it is all the predicted values) calculated from the model \mathbf{m} .

of the inversion process for each frequency. The (KM) (Kuhn 1955; Munkres 1957) algorithm is adopted to get the matching relationship of the bipartite graph for each frequency, the mode-orders of the observed values with NESMO would be dynamically given for each model \mathbf{m} during the process of model updating.

The rest of this paper is organized as follows. The PSEKM is introduced in Section 2. Then, the synthetic reconstruction tests with theoretical phase velocities of the four-layer incremental model (Model A), four-layer model with a low-velocity weak interlayer (Model B), and four-layer model with a high-velocity hard interlayer (Model C) are designed to check the performance of the novel workflow in Section 3. Also, the workflow is also tested by the observed phase velocities extracted from the synthetic seismic records of the Model B. The performance of the preconditioned steepest-descent (PSD) algorithm in the local optimization methods and the impact of the mode misidentification have also been reviewed. Then, the proposed workflow is applied to the analysis of the field surface wave data collected at two roadbeds. Finally, the role of the staging strategy of our inversion workflow is demonstrated in detail in the discussion part.

2. METHOD

2.1. The definition of the misfit function

The following notations are introduced to facilitate the description of our method. Let the set of observed phase velocities for the multimodal surface waves be v^o , where the reliable segment of the fundamental-mode surface wave is A and the others are B. And the observed phase velocities in the set B are all NESMO. The sets A and B have the following relationship:

$$\begin{cases} A \cup B = v^o, \\ A \cap B = \emptyset. \end{cases} \quad (1)$$

Here, the matrix is adopted to describe the data structure of the observed phase velocities. And the matrixes of the following synthetic reconstruction tests for the Model A (Table 1), Model B (Table 2), and Model C (Table 3) are given by Fig.1 (b), (d) and (f). And Fig. 1 (a), (c) and (e) denote the observed phase velocities for the corresponding theory model. The 0-element in the matrix means that the phase velocity measurement for the location is not implemented (Fig. 1(b)),

(d) and (f)). The set A is placed in 0-th column of the matrix, and the set B is expanded naturally in the matrix according to its own capacity (Fig. 1(b), (d) and (f)). Please note that the observed phase velocities used here are only to check the optimal ability of our proposed method, and the reasonability of the observed set is not considered. The synthetic test that is closer to the situation in the real world is given in the Section 3.2.

When the inversion is in the first stage, only the difference between the set A and that part of the predicted set is measured, and the “distance” d between them based on the L^p norm is defined as:

$$d(\mathbf{m}) = d_A(\mathbf{m}) = \sum_{i=0}^{M-1} |A(f_i) - v^{pre}(f_i, \mathbf{m})_0|^p, \quad (2)$$

where \mathbf{m} is the model occurred in the inversion process, M is the number of the observed values of the set A, f_i represents the frequency, $v^{pre}(f_i, \mathbf{m})$ denotes the predicted multimodal surface wave phase velocities at the frequency f_i calculating from the model \mathbf{m} by the multi-mode fast vector transfer algorithm (Fan & Liu 2001; Fan et al. 2007), and the subscript 0 for the $v^{pre}(f_i, \mathbf{m})$ represents the operation of getting the fundamental-mode surface wave phase velocity.

When the inversion is at the second stage, we introduce an operator d^p that minimizes the “distance” between the observed set v^o and predicted set v^{pre} calculated from the model \mathbf{m} during the inversion process based on the L^p norm.

$$d(\mathbf{m}) = d^p(v^o, v^{pre}(\mathbf{m})) = d_A(\mathbf{m}) + d_B(\mathbf{m}), \quad (3)$$

where the $d_A(\mathbf{m})$ has been defined in the formula (2), so the minimization operation is included in the calculation for the $d_B(\mathbf{m})$. And this operation acting on the observed set B can be decomposed into the minimization for the observed values at each frequency f_i , and the expression for the subproblem is:

$$d_B^i(\mathbf{m}) = \min_{j^k} \sum_{k=0}^{n_i-1} |B(f_i)_k - v^{pre}(f_i, \mathbf{m})_{j^k}|^p \quad j^k \geq b^i. \quad (4)$$

In the above formula, n_i is the number of the observed values at the frequency f_i for the set B, k is the column index, the subscript j^k denotes the operation of getting the j^k -th-mode surface wave phase velocity. And b^i is an existential identity variable for the observed values of the fundamental-mode surface wave at the frequency f_i , if it's existent, $b^i = 1$, otherwise $b^i = 0$. In fact, the subproblem described by the formula (4) is a discrete optimization problem, and it's also a many-to-many matching problem between the observed and predicted phase velocities. A weighted bipartite can be built for describing the subproblem, then the KM (Kuhn 1955; Munkres 1957) algorithm is adopted for getting the matching relationship (the identification result for the mode-order) for each frequency. The specific solution process of the subproblem and the principle of the KM algorithm are scheduled in the Section 2.3.

To reflect the average phase-velocity difference between the observed and predicted data sets, the misfit function is given as:

$$\varphi(\mathbf{m}) = \left(\frac{d(\mathbf{m})}{N} \right)^{1/p}, \quad (5)$$

where N is the number of observed phase velocities and the parameter p is set as 2 in the paper. At this time, the formula (5) reflects the root-mean-square (RMS) error between the observed and predicted data sets.

2.2. Pattern search (PS) algorithm

Geophysical inversion is essentially a process of adopting a certain algorithm or strategy to update the model \mathbf{m} to minimize the “distance” (or misfit function) between the observed and modelled data, which can also be regarded as getting a satisfactory inverted result through the sampling/travelling on the hypersurface of the misfit function values. In the past cognition, it's thought that the local minima on the hypersurface of the misfit function values affect the optimization ability and performance of the local optimization algorithms, such as the steepest-descent, quasi-Newton and preconditioned steepest-descent (PSD) algorithms. However, the work of Dauphin et al. (2014) has shown that a deeper and more profound difficulty originates from the proliferation of the saddle points, not local minima, especially the misfit function defined over continuous, high dimensional spaces. The saddle points are surrounded by the

plateaus of the high misfit function values at some dimensions that can dramatically reduce the performance of the local optimization algorithms, and give the illusory impression of the existence of a local minimum (Dauphin et al. 2014). Furthermore, the local minima or maxima are in fact rare compared to the saddle point for many high-dimensional nonconvex functions (Goodfellow et al. 2016). In other words, the probability of sampling the real local minimum when updating the model during the inversion process is very low. For the saddle point, we can always find a dimension of the model parameter to modify to optimize the misfit function value.

The pattern search (PS) is a class direct search method used to solve global optimization problems with simple boundary constraints (Conn et al. 1991). The PS method only updates at most a dimension of the model \mathbf{m} at each iteration, so it deserves to be adopted because of the feature that the hypersurface of the misfit function contains a large number of saddle points. Song et al. (2009) first introduced the PS algorithm to the inversion of surface-wave dispersion curve; however, the mode misidentification was not considered in their work.

The core of the PS algorithm can be summarized as generating multiple trial models (candidate points) around the current model \mathbf{m}_k at the k -th ($k \geq 0$) iteration and selecting the optimal model from these models. The stencil matrix \mathbf{P}_k for perturbing the current model \mathbf{m}_k is defined as:

$$\mathbf{P}_k = \mathbf{a}_k \mathbf{D}_k, \quad (6)$$

where \mathbf{a}_k is the diagonal matrix for the step-length and its diagonal element at each row defines the perturbed step-length for the model parameter at the corresponding dimension, and \mathbf{D}_k is the matrix for describing the search direction. The S-wave velocity vector \mathbf{V}_S and thickness vector \mathbf{h} of the formation are inverted (the form for the model is $\mathbf{m} = (\mathbf{V}_S, \mathbf{h})^T$ and the symbol T denotes the transpose operation) in the paper, while the P-wave velocity and density are determined, empirically. So, the number n of the variables of the inversion system is $2l-1$ when the subsurface media is parameterized l layers. At this time, the size of \mathbf{a}_k and \mathbf{D}_k would be $n \times n$ and $n \times 2n$. Then, the form of \mathbf{a}_k is expressed as:

$$\mathbf{a}_k = \begin{bmatrix} \alpha_k^{00} & & & \\ & \alpha_k^{11} & & \\ & & \ddots & \\ & & & \alpha_k^{n-2n-2} \\ & & & & \alpha_k^{n-1n-1} \end{bmatrix}. \quad (7)$$

The diagonal element at each row gives the trial step-length for the inversion parameter at the corresponding dimension at the k -th iteration. And the detailed composition of \mathbf{D}_k is:

$$\mathbf{D}_k = \mathbf{D}_0 = [\mathbf{I}, -\mathbf{I}], \quad (8)$$

where \mathbf{I} is the unit matrix with n -order and \mathbf{D}_k remains invariant for the entire inversion. Each column of the \mathbf{P}_k is used to perturb the current model \mathbf{m}_k , so $2n$ trial models would be generated for the k -th iteration, and one of them is:

$$\mathbf{m}^t = \mathbf{m}_k + \mathbf{P}_k^t. \quad (9)$$

The trial model \mathbf{m}^t is generated through utilizing the t -th column of \mathbf{P}_k to perturb the \mathbf{m}_k . Comparing the minimum misfit function value of these trial models and the misfit of the current model \mathbf{m}_k , we give the following rule for the model updatation:

$$\begin{cases} \mathbf{m}_{k+1} = \mathbf{m}^{optimal} = \mathbf{m}_k + \mathbf{P}_k^{optimal} & \text{successful prediction: if } \varphi(\mathbf{m}^{optimal}) < \varphi(\mathbf{m}_k), \\ \mathbf{m}_{k+1} = \mathbf{m}_k & \text{failed prediction: if } \varphi(\mathbf{m}^{optimal}) \geq \varphi(\mathbf{m}_k), \end{cases} \quad (10)$$

where the $\mathbf{m}^{optimal}$ is the one with the minimum misfit value among these trial models. If the prediction is successful, the step-length would be expanded or retained; otherwise, reducing the step-length to reiterate, which is:

$$\mathbf{a}_{k+1} = \begin{cases} \lambda \mathbf{a}_k & \text{successful prediction,} \\ \theta \mathbf{a}_k & \text{failed prediction,} \end{cases} \quad (11)$$

λ is the expansion factor greater than or equal to 1, and θ is the shrinkage factor less than 1 in the above formula. Once the \mathbf{m}_0 , \mathbf{a}_0 , \mathbf{D}_0 , λ , and θ are given, the iteration is started naturally according

to the formulas (6) to (11), so that the misfit value can be continuously reduced. The pseudo-code for the Pattern search (PS) is as follows:

Algorithm 1: Pattern search (PS) algorithm

Given the convergence tolerance φ_{tol} (such as $10^{-3}, 10^{-4}, 10^{-5}$, etc.) for the misfit function and that α_{tol} for the step-length matrix.

Chosen the initial model \mathbf{m}_0 , initial step-length matrix \mathbf{a}_0 , matrix \mathbf{D}_0 for the search direction, expansion factor λ and shrinkage factor θ .

for $k = 0, 1, 2, \dots, maxIter$ (maximum iteration)

if $\varphi(\mathbf{m}_k) \leq (\varphi(\mathbf{m}_0) \times \varphi_{tol})$ or $\mathbf{a}_k \leq \mathbf{a}_{tol}$

return \mathbf{m}_k ;

end (if)

$\mathbf{P}_k = \mathbf{a}_k \mathbf{D}_k$;

for $t = 0, 1, 2, \dots, 2n-1$

$\mathbf{m}^t = \mathbf{m}_k + \mathbf{P}_k^t$;

 Bounds checking and rebound for \mathbf{m}^t ;

 Calculating the misfit function value for \mathbf{m}^t and storing it;

end (for)

 Finding the model $\mathbf{m}^{optimal}$ with the smallest misfit function value from these trial models \mathbf{m}^t ;

if $\varphi(\mathbf{m}^{optimal}) < \varphi(\mathbf{m}_k)$

$\mathbf{m}_{k+1} = \mathbf{m}^{optimal}$;

$\mathbf{a}_{k+1} = \lambda \mathbf{a}_k$;

$\mathbf{D}_{k+1} = \mathbf{D}_k$;

else

$\mathbf{m}_{k+1} = \mathbf{m}_k$;

$\mathbf{a}_{k+1} = \theta \mathbf{a}_k$;

$\mathbf{D}_{k+1} = \mathbf{D}_k$;

end (if)

end (for)

return \mathbf{m}_{k+1} ;

$\mathbf{a}_k \leq \mathbf{a}_{tol}$ means that at least one diagonal element of the \mathbf{a}_k is smaller than that of the \mathbf{a}_{tol} for the corresponding location.

2.3. Kuhn–Munkres (KM) algorithm

The Kuhn–Munkres (KM) (Kuhn 1955; Munkres 1957) algorithm was proposed to solve the general assignment and transportation problems, such as choosing an optimal assignment of b men to b jobs (assuming that the numerical ratings are given for each man’s performance on each job). An optimal assignment plan can maximize the sum of the numerical ratings of b men, if all possible plans are evaluated, there are $b!$ (the symbol ‘!’ represents the factorial calculation) possibilities. The complexity of the exhaustive methods is unacceptable in most cases, and the contribution of Kuhn (1955) and Munkres (1957) is to make the problems can be solved in polynomial time. Later, this kind of assignment or transportation problem was summarized as the maximum matching problem of a weighted bipartite graph, and the KM algorithm is the most commonly-used solution.

Considering such a scenario, we calculate q phase velocity values at a certain frequency from a certain model during the iterative process, and the number of corresponding observed values with the NESMO is p_b . How should we establish a matching relationship between the observed values with NESMO and the predicted values? A weighted bipartite graph G is constructed to find the matching scheme to minimize the difference between the observed values with the NESMO and predicted values (if the measured value of the fundamental surface wave is existent at the frequency point, the number of the predicted values used here would be $q_b=q-1$ because the predicted value for the fundamental-mode surface wave would not participate in the construction of the bipartite graph, otherwise $q_b=q$) at the certain frequency. By connecting the vertices of the observed values with NESMO to the vertices of the used predicted values at the specific frequency, the weights of the edges are given by the negative distance derived from the L^p (p is set as 2) norm between the observed and predicted values, and a schematic bipartite G is shown as Fig. 2 ($p_b=2$, $q_b=3$). For the models with $p_b>q_b$, they would be rejected for evaluation in the process of model updating. The vertices of the observed and predicted values for the graph G are placed in the sets X and Y , respectively. All weights of the edges between X and Y are denoted by the weight matrix \mathbf{W} (\mathbf{W} is a matrix with the shape of 2×3 at this time, and $W(i,j)$

denotes the negative distance between the i -th observed value with NESMO and j -th used predicted value.). Then, the matching relationship of the sets X and Y is obtained by the KM algorithm. The flow of the KM algorithm (West 2020) is:

Algorithm 2: Kuhn-Munkres (KM) algorithm

Step 1: Let the feasible vertex label of the vertex X_i be $a[i]$ and that of the vertex Y_j be $b[j]$. Initially, $a[i]$ is set as the maximum weight of the edge associated with X_i , and $b[j]=0$, ensuring that $a[i]+b[j]\geq W(i,j)$ holds.

Step 2: Using the Hungarian method (Kuhn 1955) to find a maximum match O from the graph G, if O is a complete match (the number of edges in graph O is p_b at this time), the match O is returned as the maximum weight match. Otherwise, let Q be a set of the vertex cover with size $|O|$ in G. Assuming $R=Q\cap X$, $T=Q\cap Y$. Then, the term of ε is calculated by:

$$\varepsilon = \min \{a[i] + b[j] - W(i, j) : X_i \in X - R, Y_j \in Y - T\}. \quad (12)$$

For $X_i \in X - R$, subtracting ε from $a[i]$; for $Y_j \in T$, adding ε to $b[j]$. After this step, a new equal subgraph ($a[i]+b[j]=W(i, j)$ is always maintained in the subgraph) is formed, and repeat the iteration process.

Finally, the match relationship of the vertex X_1 matches Y_2 and the vertex X_2 matches Y_3 is obtained. The detailed calculation process for the bipartite G is shown by the Appendix B.

3. SYNTHESIZED TESTS

3.1. Reconstruction tests of theory model with error-free data

To check the performance of the proposed workflow, the reconstruction tests of three typical four-layer models (Models A, B, and C) are designed, whose properties are described in Tables 1, 2 and 3. The observed phase velocities of the corresponding model during the tests are calculated by the multimodal fast vector transfer algorithm (Fan & Liu 2001; Fan et al. 2007). Except for the phase velocities of the fundamental-mode, we assume that the mode-orders of the other observed values are not directly available (NESMO). The observed phase velocities and

data structures for the corresponding models have been shown by Fig. 1. The inversion is divided into two stages. First stage: We set a half-space or an incremental S-wave velocity model with equal layer thickness as the initial model to invert the phase velocities of the fundamental-mode surface wave through the PS algorithm. Second stage: The inverted result of the first stage is taken as the initial model of the second stage and the PSEKM method is adopted to invert all the observed phase velocities. The mode-orders of the observed values with NESMO at each frequency are dynamically predicted for each model \mathbf{m} occurred in the inversion process. The S-wave velocity (V_S) and thickness (h) of the formation would be updated during the inversion process, and the ratio of P-wave velocity (V_P) to S-wave velocity and the density (ρ) remain known for the inversion of synthetic data. To quantitatively evaluate quality of the inversion result, we give the definition of the relative error err between the inverted result and true model, showing by:

$$err = \frac{1}{n} \sum_{i=0}^{n-1} \left| \frac{m_i^{inv} - m_i}{m_i} \right|, \quad (13)$$

where n is the number of the inversion parameters, and the i -th parameter of the true model and inverted result are denoted by m_i and m_i^{inv} , respectively.

3.1.1. Reconstruction test for the Model A

The parameters for the Model A are shown in Table 1, which is a four-layer model with the incremental velocity structure, and the last column of the table gives the boundary constraint information about the S-wave velocity and thickness for each layer during the inversion process. According to the previous description, the initial model is set as a four-layer incremental model with equal thickness of each layer (here, the depth information of the half-space is used), and the construction method for the initial model is described in detail in the Appendix A. The initial step-lengths of the S-wave velocity and layer thickness are chosen as 1 and 0.02, respectively, the expansion factor λ is set as 1.2, the shrinkage factor θ is taken as 0.5, and the inversion would be terminated when the number of the iterations reaches to 500 or the step-lengths for the S-wave velocity and layer thickness meet the tolerance (the parameter for the S-wave velocity is less than 0.001 and that for the layer thickness is smaller than 0.0001). The above hyperparameters for the inversion setting are remained for the synthetic tests in the Section 3.1. Fig. 3 shows the

evaluated results about the model reconstruction. After the first-stage of inversion, the observed phase velocities of the set A have been almost faultlessly fitted, what's more, the mode-orders of the observed values with the NESMO are also provided according to the predicted values of the inverted result (Fig. 3 (a)). This shows that the fundamental-mode surface wave with a sufficiently broad-band can predict the mode-orders information about the observed phase velocity values with NESMO (Fig. 3 (a)). However, the manual judgement on the mode-orders is inconvenient and time-consuming. The function of the KM algorithm is to enable the mode-order information of the data points with the NESMO to be dynamically determined during the inversion, so that the field data containing the mode losses and aliases is treated automatically. The predicted values calculated from the inversion result of the second stage almost perfectly match all the observed values, and there is also no difference between the predicted and analytical values (Fig.3 (b)). It can be seen that the initial misfit value of the second stage is greater than the final misfit value of the first stage (Fig. 3 (c)), which indicates that the observed phase velocities of the set B store the extra information about the model. After the inversion of two stages, the model has been reconstructed with extremely high accuracy (Fig. 3 (d)). The relative error between the inverted result of the first stage and the true model is 5.29%, and that of the second stage is 0.63%. Here, the surface waves of the NESMO play a role in the modification of the model determined by the fundamental-mode surface wave, making the inverted result closer to the true model (Fig. 3 (d)).

3.1.2. Reconstruction test for the Model B

The specific parameters for the Model B and the boundary constraints for the S-wave velocity and layer thickness of the inversion are given in the Table 2. The initial model of the inversion is set as the homogeneous half-space model with the S-wave velocity at 310 m/s (Fig.4 (d)), and the established method is shown in the Appendix A. After the first stage, the observed fundamental-mode surface wave has been nicely fitted, but the predicted values of the higher-mode surface waves calculated from the inverted result are still slightly different from the observed values of set B (Fig. 4 (a)). This also indicates that the observed phase velocities in the set B store the different information about the underground model compared with the fundamental-mode surface wave (Fig. 4 (a) and (c)). There is no visual difference between all the observed and predicted values, and all analytical phase velocity values are also accurately

matched after the second stage (Fig. 4 (b)). The parameters for describing the true model have been retrieved with extremely high precision (Fig. 4 (d)), and the relative errors between the inverted results and the true model are reduced from 8.42% to 0.57% due to the addition of surface wave observation with NESMO. Even, the inversion result has coincided exactly with the true model, visually (Fig. 4 (d)).

3.1.3. Reconstruction test for the Model C

Table 3 shows the detailed properties of the Model C and the boundary constraint information of the related parameters during the inversion. The initial model for the first stage of the inversion is also taken as the homogeneous half-space model (Fig. 5 (d)), and the Appendix A gives the construction method. The observed values used at each stage of the inversion are always accurately fitted by our presented method (Fig. 5 (a) and (b)). The relative errors between the reconstructed results and the true model have also been reduced from 10.51% to 3.26% due to the staging strategy and the addition of the surface wave observation with NESMO (Fig. 5 (d)). And the observed phase velocities in the set B store the extra information about the model, which has been verified again (Fig. 5 (a), (c) and (d)).

3.2. Reconstruction tests of the Model B with measured data

3.2.1. Reconstruction test for the Model B with proposed workflow

The observed values selected in the above synthetic tests are all error-free and not considered the reasonability. To simulate the surface wave phase velocity measurement in real world, we synthesize the Z-component records for the Model B through the discrete wavenumber method (Bouchon 2003; Kausel, 2018; Zhang et al. 2020), the 30 Hz Ricker wavelet with the delaying-time of 16 ms is loaded on the free surface as a vertical force source, and Fig. 6 (a) shows the synthesized records. The nearest offset of the records is 7 m, 60 traces are recorded with a receiver-spacing of 1 m, and the number of the sampling points is 2048 with the sampling rate at 1000 Hz. Only the signals within 0.7 s are shown for clarity (Fig. 6 (a) and (b)). Taking the first trace of the synthetic records as the reference (virtual source), the cross-correlation function (CCF) gather shown by Fig. 6 (b) is obtained through the cross-correlation. Then, the F-

J transform is used to generate the dispersion energy image (Fig. 6 (c)) from the CCF gather, and the specific formula for the calculation of the F-J spectrum E (Wang et al. 2019) is:

$$E(\omega, k_o) = \int_0^{+\infty} C(\omega, r) r J_0(k_o r) dr, \quad (14)$$

where the $C(\omega, r)$ is the spectrum of the CCF at the circular frequency ω , r represents the distance between the virtual source and receiver-point, k_o is the circular wavenumber, and J_0 is zero-order Bessel function of the first kind. Both the numerical integration and domain-conversion from $\omega-k_o$ to $\omega-v$ (phase velocity) domain about the above formula have been clearly expressed in the reports of Wang et al. (2019) and Li et al. (2021). It needs to be indicated that we use the CCF gather rather than the raw surface wave records to complete the dispersion measurement to remove the adverse effect of the initial phase at each frequency of the source wavelet.

The real part of the formula (14) (F-J spectrum) is shown as the dispersion energy image, whose negative elements are zeroed (Fig. 6 (c)). The black dotted lines are the picked dispersion curves for the multimodal surface waves (Fig. 6 (c)), and their mode-orders are 0-th, 2-nd and 5-th, respectively (Fig. 6 (d)). The observed set A stores the phase velocities for the fundamental-mode surface wave and the phase velocities in the set B are all NESMO for the inversion at the second stage. The observed phase velocities belonged to the 5-th mode have a considerable deviation from the analytical values, which may be caused by the end effect during the F-J transform.

The initial model for the inversion at the first stage is also taken as the half-space model with the S-wave velocity at 322 m/s, and the detection depth is determined to be the 0.75 times the mean wavelength of the fundamental-mode surface wave (Fig. 7 (d)). The density and ratio between the V_P and V_S are considered known during the inversion process. The boundary constraints for the inverted parameters remain the same as the synthetic test in the Section 3.1.2 (Table 2). The expansion factor λ is set as 1.0, the shrinkage factor θ is also taken as 0.5, the initial step-lengths for the S-wave velocity and formation thickness are set as 5 and 0.1, respectively, the tolerance parameters for them are given as 0.1 and 0.001, and the maximum iterations is still set as 500 for the inversion. After the inversion of the first stage, the key feature

of the stratigraphic structure containing a low-velocity layer has been revealed (Fig. 7 (d)). And the relative error between the inverted result and true model is 22.24%. The mode-order information of the observed phase velocities in the set B can also be correctly inferred (Fig. 7 (a)), and the mode misidentification doesn't appear at this time, which may benefit from the special way of model updating of the PS algorithm. Even so, the suggested workflow avoids the bored manual identification for the surface-wave mode-orders. Moreover, the classical local optimization methods are more prone to run into the mode misidentification, which would be checked by the following numerical test in the Section 3.2.2. After the inversion of the second stage, all observed phase velocities are nicely fitted with the right mode-orders (Fig. 6 (d) and Fig. 7 (b)). The performance may be case-dependent, but for the field data, we provide an ex-post evaluation scheme to analyze the credibility of the inversion result (Fig. 9). The initial misfit value at the second stage is also greater than the final misfit of the first stage, showing that the observed phase velocities in the set B store the different information about the true model (Fig. 7 (c)), again. The inverted result of the second stage gives more accurate burial depth and layer thickness for the critical weak layer (Fig. 7 (d)), and the relative error is reduced to 21%. More importantly, the addition of the observed phase velocities with NESMO stabilizes the distribution of the relative errors between the inversion result and true model about the parameters in all dimensions (Fig. 7 (e)). Specifically, the maximum relative error of the parameters between the inverted result of the first stage and true model is 81.2%, and that of the second stage is only 49.33%.

3.2.2. The performance of the PSD and mode misidentification

To verify the advantages of the proposed workflow, we select the PSD algorithm belonged to the local optimization method to invert the observed phase velocities of the Model B. And the detailed formula and robust implementation for the PSD algorithm are arranged in the Appendix C. The entire inversion is also divided into two stages. Only the fundamental-mode surface wave observations are inverted for the first stage. Then, we calculate the predicted multimodal dispersion curves of the inverted result of the first stage, and the mode-orders of the higher-mode surface wave are manually judged according to the closeness between the predicted values and dispersion energy. So, the inversion of the second stage is carried out naturally once the mode-

orders of the observed phase velocities are artificially designated. The bound constraints and initial model remain the consistent with the above test in the Section 3.2.1.

After the first stage of the inversion, the fundamental-mode observations have been fitted well with very small RMS value (Fig. 8 (a) and 8 (c)), and the critical feature of the stratigraphic structure containing a low-velocity layer has also been revealed (Fig. 8 (d)). The relative error between the inverted result of the first stage and true model is 29.82 %, however, that for the proposed workflow is only 22.24%. Regrettably, the observed phase velocities belonging to the 5-th surface-wave mode are misjudged as the 6-th mode. After the second-stage inversion with wrong mode identification, the entire observed phase velocities are still fitted properly (Fig. 8 (b)). Scilicet, from the perspective of data fitting, the inversion result must be satisfactory. Unfortunately, the key feature of the stratigraphic structure including a weak interlayer is blurred (Fig. 8 (d)). The relative error between the inverted result of the second stage and true model is dramatically increased to 45.48%, while that for the proposed workflow is reduced to 21%, which has clearly verified the stability and advantages of the novel workflow and method. This also indicates that the wrong mode identification has a negative impact on the description of the subsurface media.

4. FIELD DATA CASES

4.1. The observed system and analysis workflow for the field data

The field data sets are collected on two roadbeds of the Shen-Bai high-speed railway line, China. The 4.5 Hz receivers are chosen to record the Z-component signals, and there are 24 geophones deployed on the free surface with 1 m interval for each survey point. The recorded Rayleigh-wave is excited by a sledgehammer with the nearest offset at 7 m. The sampling rate of the signal is set as 2000 Hz, and the number of the sampling points is 4096.

According to the above description, we give the following analysis workflow (Fig. 9) for the field data. The analysis workflow may be divided into three steps, which are:

Step 1, Dispersion measurement: Generating the CCF gather from the field surface wave data with the first trace as the reference through the cross-correlation. The purpose of the cross-correlation operation is to remove the initial phase at each frequency of the source, and the

deconvolution-based methods are also strongly recommended (Wapenaar et al. 2011). Then, the F-J transform is used to extract the dispersive spectrum from the CCF gather. The measured values for the multimodal surface waves are picked from the dispersion image, and dividing the sub-sets A and B of all observed phase velocities. The observed set A would store the phase velocities for the fundamental-mode surface wave, and the observed values in set B are all NESMO.

Step 2, Dispersion inversion: The whole inversion is divided into two stages. Constructing the half-space or incremental initial model by the method in the Appendix A. Then, only the observed phase velocities in the set A are inverted for the first stage by the PS algorithm, adding all the observations to the inversion system and the inverted result of the first stage is set as the initial model for the inversion at the second stage.

Step 3, Credibility analysis: The 1D conditional probability density distribution (CPDD) of the inversion parameters at each dimension is calculated for analyzing the credibility of the inverted result (Song et al. 2020). Also, synthesizing the waveform of the inverted result by the discrete wavenumber method for further credibility analysis.

4.2. Roadbed 1

The normalized seismograms of the raw Rayleigh-wave records are shown as Fig. 10 (a), and the CCF gather (Fig. 10 (b)) is generated with the first trace as the reference trace (virtual source). The red stars of Fig. 10 (a) and (b) point out the horizontal location for the coring operation, and a bad trace is also formed at the location, possibly because the previous drilling had broken the free-surface coupling. No extra operations are applied to the bad trace during the calculation for the dispersive spectrums due to the large amplitude characteristics of surface waves. And Table 4 gives the detailed description for the borehole data and S-wave velocity range of each formation for the roadbed 1 deduced from *Code for seismic design of railway engineering* of China. The dispersive spectrums calculated by the 2D Fourier transform and phase-shift have almost the same data distribution and resolution (Fig. 10 (c) and (d)), and the nearly consistent observed phase velocities for the multimodal surface waves must be picked from the dispersive spectrum according to the peaks of the image. The similar measurement can also be attained for the HRLRT (Fig. 10 (e)) with higher resolution. The performance of the F-J transform is surprising, which provides the observation for the extra surface-wave mode (Fig. 10

(c), (d), (e) and (f)). Also, its resolution is basically the same as the HRLRT. This may benefit from the cylindrical wave assumption of the F-J transform that is closer to the wave propagation mechanism in real world, while other methods are based on the plane wave diffusion. Furthermore, more observations extracted by the F-J transform would further reduce the multi-solution of the inversion. Certainly, it's also very difficult to judge the mode-order of this extra observation for the surface waves directly from the dispersive spectrum (Fig. 10 (f)), that's why we propose the modern inversion workflow in the article. The green dotted line and red dotted lines on the Fig. 10 (f) give the observed phase velocities with 1 Hz interval for the fundamental-mode surface wave and NESMO. Also, the fundamental-mode observations are stored in the set A, and the others are placed in the set B.

Once the observed phase velocities for the multimodal surface waves are extracted, the proposed inversion workflow are proceeded. Utilizing the method in the Appendix A, a constant S-wave velocity model with a 4-layer structure is constructed as the initial model for the inversion at the first stage, and the detection depth is determined as the mean wavelength of the fundamental-mode surface wave with a priori (Fig. 11 (d)). The density and the ratio V_p/V_s remain invariant for each layer during the whole inversion, and they are set as 2000 kg/m^3 and 2.45, respectively. The searching range of the S-wave velocity for each layer is set as 50~1000 m/s, and that for the layer thickness is 0.5~10 m. The expansion factor λ is set as 1.0, the shrinkage factor θ is taken as 0.5, the initial step-lengths for the S-wave velocity and formation thickness are set as 5 and 0.5, respectively, the tolerance parameters for them are given as 0.1 and 0.01, and the inversion for each stage would be automatically terminated when the number of iterations exceeds 500.

After the first-stage inversion, the observed values of the fundamental-mode surface wave have been fitted nicely with a very small RMS value (Fig. 11 (a) and (c)). Moreover, the observed phase velocities with lower values in the set B can also be properly matched (Fig. 11(a)). The RMS between the fundamental-mode surface wave phase velocities of the inverted result at the first stage and that in the set A is only 2.04 m/s, and the inversion result is completely acceptable when only the fundamental-mode surface wave observations are considered in the data domain. However, the dispersive observations with higher phase velocities in the set B are not fitted by the phase velocities of the multimodal surface wave for the inverted

result at the first stage (Fig. 11 (a)). Furthermore, the revealed half-space S-wave velocity is only about 400 m/s, while the bedrock is occurred after the depth of 9.10 m (Table 4). The S-wave velocity of the bedrock composed of weakly weathered amphibolite is not too low, whose range is 500~1000 m/s, and it is also 750 m/s when taking the median value. And the RMS value rises to 84.05 m/s (Fig. 11 (c)) when all observations for surface waves are considered in the data domain. That's to say, the inverted result for the first stage performs well on the observations for the fundamental-mode surface wave, however, perform poorly on the whole observed dataset (Fig. 11 (a) and (c)). This again shows that the observed phase velocities in the set B store the additional information about the S-wave velocity structure of the subsurface medium.

After the inversion at the second stage, all the observed values for the surface waves have been fitted almost perfectly with the RMS value at 4.90 m/s (Fig. 11 (b) and (c)). The possible mode loss of surface waves has been revealed from the result of the second stage (Fig. 11 (b)). The set B may store the observed phase velocities for the 1-st and 6-th surface-wave mode (Fig. 11 (b)). The inverted result of the first stage is significantly different from that of the second stage (Fig. 11 (d)). And the inversion result of the first stage shows an increasing trend for the S-wave velocity structure with a smaller half-space velocity. However, the inverted S-wave velocity structure of the second stage includes the higher half-space velocity and deeper detection depth (Fig. 11 (d)). From the fitting of all surface wave observed phase velocities, the inversion result at the second stage is more preferable (Fig. 11 (a) and (b)). Especially after the inversion at the second stage, the more reasonable half-space S-wave velocity is revealed (Fig. 11 (d) and Table 4) due to the extra observations of surface waves provided by the F-J transform. So, it's easy to infer that the dispersion measurement of this case based on the plane wave assumption can't perceive the more realistic half-space S-wave velocity (Fig. 10 (c), (d) and (e)).

When the borehole data is not available, how to semi-quantitatively evaluate the reliability of the parameters for the final inversion result at each dimension in the data domain? We refer to the practice of Song et al. (2020) to calculate the 1D conditional probability density distribution (CPDD) of the inversion parameters for the second stage to evaluate the credibility at each dimension. The 1D CPDD of the inversion parameters at each dimension is shown as Fig. 12, which of each dimension is obtained through fixing the parameters at other dimensions of the inverted result, and only traversing the parameter of this dimension at its searching range to

calculate the corresponding conditional probability. Also, the conditional probability value is inversely proportional to the misfit function value. The red pentagrams mark the inversion parameters of the inversion result at each dimension (Fig. 12), which always correspond exactly to the peak of the distribution. This provides important evidence that the result is credible, and the confidence interval of the inversion parameters are very narrow, except for the half-space S-wave velocity (Fig. 12), which may reflect that the dispersive observations have different sensitivities to different parameters.

In addition to the 1D CPDD, we also synthesize the waveform of the inversion result at the second stage by the discrete wavenumber method for further analysis, and the first trace of the field data is selected as the source wavelet due to its spectral similarity to the true wavelet. The synthesized seismograms are shown as Fig. 13 (a), and its CCF gather is generated with the first trace as the virtual source (Fig. 13 (b)). The predicted and observed CCF waveforms have been matched with high precision (Fig. 13 (b)) except for the bad trace at the offset 20 m, which provides another evidence for the credibility. The high-velocity dispersion energy at the high-frequency band may be inferred as the guided P-wave (Fig. 10 (f)) when the half-space S-wave velocity range is unknown. For this, we use the frequency(f)-wavenumber(k) filter to separate the required part from the observed and predicted CCF waveforms. The observed CCF waveforms with phase velocity less than 400 m/s is fitted with high accuracy by that part of the predicted data (Fig. 13 (c)), and the matching degree is still acceptable between the observed and predicted CCFs with the phase velocity greater than 400 m/s (Fig. 13 (d)). Then, the corresponding dispersion images are generated by the F-J transform (Fig. 13 (e) and (f)). Compared with the field data, the fundamental-mode surface wave is more continuous, so the dispersion energy about the first-order surface wave does not appear (Fig. 13 (e)), which may be caused by the inaccurate source wavelet. The high-velocity dispersion energy at the high-frequency band is still generated for the corresponding part of predicted data, but it needs to be pointed out that the energy is weak for the certain band (Fig. 13 (f)). There are many reasons for this, it could be the inaccurate seismic wavelet, the model incompatibility between the field data and the predicted data (such as the observed data contains the attenuation and horizontal heterogeneity, while the synthetic data does not) and so on.

4.3. Roadbed 2

The same workflow is adopted to get the S-wave velocity structure of the roadbed 2. The parameters of the inversion setting remain the same as those for the roadbed 1, except that the searching range of the S-wave velocity is changed to 50~1500 m/s. The half-space initial model with a 4-layer structure is also built by the method in the Appendix A, and its detection depth is determined as the mean wavelength of the fundamental-mode surface wave with a priori. The density and ratio value of V_p/V_s are also set as 2000 kg/m³ and 2.45 during the whole inversion process, respectively. After the inversion of the first stage, the observed phase velocities for the fundamental-mode surface waves have been fitted nicely with very small RMS value (Fig. 14 (a) and (c)). But the observations in the set B with higher phase velocities aren't matched by the multimodal surface wave phase velocities of the inverted result at the first stage. When the inverted result is evaluated on the all observed phase velocities, the RMS increases to about 160 m/s. This again shows that the observations in the set B store the additional information about the model of the underground media. All phase velocity observations are accurately fitted with RMS value at 6.55 m/s after the inversion at the second stage (Fig. 14 (b) and (c)). Also, it's easy to find that the mode loss is existent for the roadbed 2. The inversion result of the first and second stages are both incremental, but the result for the second stage has larger S-wave velocity values and detection depth. This indicates that the higher-mode surface waves store the deeper S-wave velocity structure information compared with the fundamental-mode surface wave. That's to say, the higher-mode surface waves can "see" deeper than the fundamental-mode surface wave with the identical wavelength, and the similar conclusion is also given in the work of Xia et al. (2003).

Then, the 1D CPDD for the parameters of the final inversion result at each dimension is given to evaluate the credibility (Fig. 15). The inversion parameter for each dimension always corresponds to the peak of its 1D CPDD, accurately (Fig. 15), which provides evidence for the reliability of the inverted result. The confidence interval for the half-space S-wave velocity and thickness of the third layer is slight wide (Fig. 15), which is the manifestation of the dispersion observations having different sensitivities to the different parameters about the subsurface media.

Similarly, we synthesize the waveforms of the inversion result at the second stage by the discrete wavenumber method, and the first trace of the field data is also chosen as the source wavelet. Then, the predicted CCF gather (Fig. 16 (b)) is obtained with the first trace of the synthetic seismograms (Fig. 16 (a)) as the reference, and the observed CCFs are still properly fitted by the predicted ones (Fig. 16 (a) and (b)). Taking the phase velocity of 400 m/s as the bound, we use the f - k filter to separate out the parts of phase velocity less and greater than the bound from the observed and predicted CCF waveforms, respectively, showing by the Fig. 16 (c) and (d). The observed CCF waveforms after separation are properly fitted by the predicted ones, especially the part of the phase velocity greater than 400 m/s (Fig. 16 (c) and (d)). The fundamental-mode surface wave dispersion energy of the predicted data with smaller phase velocity is more continuous, and the weak first-order dispersion energy is also occurred (Fig. 16 (e)). The high-velocity dispersion energy at the high-frequency band is also generated for the predicted CCF waveforms with greater phase velocity, and it is in good agreement with the corresponding observed phase velocities (Fig. 16 (f)). In conclusion, considering the 1D CPDD, waveform fitting and dispersion energy comparison, we believe that the inverted result at the second stage for roadbed 2 are very credible.

5. DISCUSSION

The natural benefit of using the staging strategy is that the inversion is started by simply constructing a half-space initial model, and it's easy to give such an initial guess based on the method in the Appendix A. If the staging strategy isn't adopted in the workflow for inversion, the half-space model cannot be set as the initial model of the inversion for multimodal dispersion curves. To probe the necessity of the staging strategy, we compare the reconstruction performance of the staging and non-staging strategy for the Model B (the observed data used here is consistent with the example in 3.1.2). We give an incremental model with an overly constant velocity layer as the initial model (Fig. 17 (b)). The inversion with the non-staging strategy is automatically terminated after about 70 iterations (Fig. 17 (a)), giving a poor reconstruction result (Fig. 17 (b)). The staging one can continue to iterate more than 300 times (Fig. 17 (a)), finally giving the inversion result almost without error, and the inversion result is visually identical to the true model (Fig. 17 (b)). The staging strategy uses the nature of the misfit function of two stages, and a game effect may be produced in the process of the model

reconstruction. Adding the observed data to the inversion system with hierarchical strategy can enhance the performance of the inversion method, this conclusion has been confirmed by many applications, such as the multiscale strategy of full-waveform inversion (Bunks et al. 1995; Yan et al. 2020).

6. CONCLUSION

For the mode misidentification caused by the mode loss in the inversion of the multimodal surface waves, we propose a modern inversion workflow with the staging strategy and PSEKM algorithm to deal with the problem. Compared with the traditional inversion method, the novel algorithm allows the existence of the observed phase velocities with NESMO in the inversion system, which weakens the conditions for the inversion to proceed, avoids the time-consuming manual mode identification and is more general in practice. The synthetic reconstruction tests of the theoretical models with the analytical or measured phase velocities have proved that the novel workflow can perceive the mode-orders of the observed phase velocities with the NESMO, and the predicted values calculated from the inversion result can always accurately match the observed values. The shortcomings and easier mode misjudgment of the local optimization method are also commented. Then, we use the proposed workflow to analyze the field data sets of the two roadbeds, and the performance is still excellent. Certainly, our presented workflow is also suitable for the analysis of the passive surface waves.

ACKNOWLEDGEMENTS

This work is supported by National Natural Science Foundation of China (Grant No. 41790465), Shenzhen Key Laboratory of Deep Offshore Oil and Gas Exploration Technology (Grant No. ZDSYS20190902093007855), and National Natural Science Foundation of China (Grant No. U1901602). We express the deepest gratitude to the anonymous reviewer, whose professional comments and thoughtful suggestions have greatly improved this article.

DATA AVAILABILITY

The surface wave data used in the study are available upon request.

REFERENCES

- Aki, K., & Richards, P. G. 1980. Quantitative seismology, Vol. II, W.H. Freeman and Company.
- Aster, R., Borchers, B. & Thurber, C. 2018. Parameter estimation and inverse problems, Vol. III, Elsevier.
- Bouchon, M. 2003. A Review of the Discrete Wavenumber Method, *Pure and Applied Geophysics*, **160**, 445-465, <https://doi.org/10.1007/PL00012545>.
- Bunks, C., Saleck, F. & Chavent, G. 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**(5), 1457-1473, <https://doi.org/10.1190/1.1443880>.
- Cai, W., Song, X., Yuan, S. & Hu, Y. 2017. A new misfit function for Multimode Dispersion Curve Inversion of Rayleigh Waves, *Earth Science*. (in Chinese), **42**(9), 1608-1622, <https://doi.org/10.3799/dqkx.2017.531>.
- Conn, A., Gould, N. & Toint, P. 1991. A Globally Convergent Augmented Lagrangian Algorithm for Optimization with General Constraints and Simple Bounds, *SIAM Journal On Numerical Analysis*, **28**(2), 545-572, <https://doi.org/10.1137/0728030>.
- Dauphin, Y., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S., Bengio, Y. 2014. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization, *NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2, 2933-2941, <https://dl.acm.org/doi/abs/10.5555/2969033.2969154>.
- Ernst, F. 2008. Multi-mode inversion for P-wave velocity and thick near-surface layers, in *Proceedings of the Near Surface 2008 14th European Meeting of Environmental and Engineering Geophysics*, Kraków, Poland, <https://doi.org/10.3997/2214-4609.20146236>.
- Fan, Y. & Liu, J. 2001. Research on the dispersion of Rayleigh waves in multilayered media, *Journal of HarBin Institute of Technology*. (in Chinese), **33**(5), 577-581, <https://doi.org/10.3321/j.issn:0367-6234.2001.05.001>.
- Fan, Y., Chen, X., Liu, X., Liu, J. & Chen, X. 2007. Approximate decomposition of the dispersion equation at high frequencies and the number of multimodes for Rayleigh waves, *Chinese Journal of Geophysics*. (in Chinese), **50**(1), 233-239, <https://doi.org/10.3321/j.issn:0001-5733.2007.01.029>.
- Forbriger, T. 2003. Inversion of shallow-seismic wavefields: I. Wavefield transformation, *Geophysical Journal International*, **153**(3), 719-734, <https://doi.org/10.1046/j.1365-246X.2003.01929.x>.
- Gabriels, P., Snider, R. & Nolet, G. 1987. In situ measurements of shear-wave velocity in sediments with higher-mode Rayleigh waves, *Geophysical Prospecting*, **35**(2), 187-196, <https://doi.org/10.1111/j.1365-2478.1987.tb00812.x>.
- Gao, L., Xia, J. & Pan, Y. 2014. Misidentification caused by leaky surface wave in high-frequency surface wave method, *Geophysical Journal International*, **199**(3), 1452-1462, <https://doi.org/10.1093/gji/ggu337>.
- Goodfellow, I., Bengio, Y. & Courville, A. 2016. Deep learning. <https://www.deeplearningbook.org>.
- Hadiouche, O., Jobert, N., & Romanowicz, B. 1986. First two-station results for long-period surface waves velocity from the Geoscope Stations in Africa, *Geophysical Research Letters*, **13**(6), 547-550, <https://doi.org/10.1029/GL013i006p00547>.
- Kausel, E. 2018. Generalized stiffness matrix method for layer soils, *Soil Dynamics and Earthquake Engineering*, **115**, 663-672, <https://doi.org/10.1016/j.soildyn.2018.09.003>.
- Kuhn, H. 1955. The Hungarian method for the assignment problem, *Naval Research Logistics*, **2**, 83-97, <https://doi.org/10.1002/nav.3800020109>.
- Li, J., Feng, Z. & Schuster, G. 2017. Wave-equation dispersion inversion, *Geophysical Journal International*, **208**(3), 1567-1578, <https://doi.org/10.1093/gji/ggw465>.

- Li, X., Chen, X. & Yang, Z., Wang, B. & Yang, F. 2020. Application of high-order surface waves in shallow exploration: An example of the Suzhou river, Shanghai, Chinese Journal of Geophysics. (in Chinese), **63**(1), 247-255, <https://doi.org/10.6038/cjg2020N0202>.
- Li, Z., Zhou, J., Wu, G., Wang, J., Zhang, G., Dong, S., Pan, L., Yang, Z., Gao, L., Ma, Q., Ren, H., & Chen, X. 2021. CC - FJpy: A Python Package for Extracting Overtone Surface - Wave Dispersion from Seismic Ambient - Noise Cross Correlation, Seismological Research Letters, **92**(5), 3179–3186, <https://doi.org/10.1785/0220210042>.
- Li, Z., Shi, C., Ren, H., & Chen, X. 2022. Multiple leaking mode dispersion observations and applications from ambient noise cross-correlation in Oklahoma, Geophysical Research Letters, **49**, e2021GL096032, <https://doi.org/10.1029/2021GL096032>.
- Lu, L., Wang, C., & Zhang, B. 2007. Inversion of multimode Rayleigh waves in the presence of a low-velocity layer: numerical and laboratory study, Geophysical Journal International, **168**, 1235-1246, <https://doi.org/10.1111/j.1365-246X.2006.03258.x>.
- Luo, Y., Xia, J., Miller, R.D., Xu, Y., Liu, J. & Liu, Q. 2008. Rayleigh-wave dispersive energy imaging using a high-resolution linear Radon transform, Pure and Applied Geophysics, **165**(5), 903-922, <https://doi.org/10.1007/s00024-008-0338-4>.
- Maraschini, M., Ernst, F., Foti, S. & Socco, L.V. 2010a. A new misfit function for multimodal inversion of surface waves, Geophysics, **75**(4), G31-G43, <https://doi.org/10.1190/1.3436539>.
- Maraschini, M. & Foti, S. 2010b. A Monte Carlo multimodal inversion of surface waves, Geophysical Journal International, **182**, 1557-1566, <https://doi.org/10.1111/j.1365-246X.2010.04703.x>.
- Mari, J. L. 1984. Estimation of static corrections for shear-wave profiling using the dispersion properties of Love waves, Geophysics, **49**(8), 1169–1179, <https://doi.org/10.1190/1.1441746>.
- McMechan, G.A. & Yedlin, M.J. 1981. Analysis of dispersive waves by wavefield transformation, Geophysics, **46**(6), 869-874, <https://doi.org/10.1190/1.1441225>.
- Munkres, J. 1957. Algorithms for the Assignment and Transportation Problems, Journal of the Society for Industrial and Applied Mathematics, **5**(1), 32-38, <https://doi.org/10.1137/0105003>.
- Nocedal, J., & Wright, S. 2006. Numerical Optimization (Second Edition), Springer.
- O' Neil, A., & Matsuoka, T. 2005. Dominant Higher Surface-wave Modes and Possible Inversion Pitfalls, Journal of Environmental and Engineering Geophysics, **10**(2), 185-201, <https://doi.org/10.2113/JEEG10.2.185>.
- Pan, J., Li, Y., Wu, Q. & Ding, Z. 2015. Phase velocity maps of Rayleigh waves in the southeast Tibetan plateau. Chinese Journal of Geophysics. (in Chinese), **58**(11), 3993-4006, <https://doi.org/10.6038/cjg20151109>.
- Pan, L., Chen, X., Wang, J., Yang, Z., & Zhang, D. 2019. Sensitivity analysis of dispersion curves of Rayleigh waves with fundamental and higher modes, Geophysical Journal International, **216** (2), 1276-1303, <https://doi.org/10.1093/gji/ggy479>.
- Papadopoulou, M., Da Col, F., Mi, B., Bäckström, E., Marsden, P., Brodic, B., Malehmir, A. & Socco, L.V. 2020. Surface-wave analysis for static corrections in mineral exploration: A case study from central Sweden, Geophysical Prospecting, **68**, 214-231, <https://doi.org/10.1111/1365-2478.12895>.
- Park, C.B., Miller, R.D. & Xia, J. 1998. Imaging dispersion curves of surface waves on multi-channel record, SEG Technical Program Expanded Abstracts, 1377-1380, <https://doi.org/10.1190/1.1820161>.
- Park, C.B., Miller, R.D. & Xia, J. 1999. Multichannel analysis of surface waves, Geophysics, **64**(3), 800-808, <https://doi.org/10.1190/1.1444590>.
- Press, W.H., Teukosky, S.A., Vetterling, W.T. & Flannery, B.P. 1992. Numerical recipes in C, Cambridge Univ. Press.
- Ryden, N. & Park, C.B. 2006. Fast simulated annealing inversion of surface waves on pavement using phase-velocity spectra, Geophysics, **71**(4), R49-R58, <https://doi.org/10.1190/1.2204964>.

- Socco, L.V., Foti, S. & Boiero D. 2010. Surface-wave analysis for building near-surface velocity models-Established approaches and new perspectives, *Geophysics*, **75**(5), 75A83-75A102, <https://doi.org/10.1190/1.3479491>.
- Song, X., Li, D., Gu, H., Liao, Y., & Ren, D. 2009. Insights into performance of pattern search algorithms for high-frequency surface wave analysis, *Computer & Geosciences*, **35**(8), 1603-1619, <https://doi.org/10.1016/j.cageo.2009.01.007>.
- Song, Z., Zeng, X., Xu, S., Hu, J., Sun, T., & Wang, B. 2020. Distributed Acoustic Sensing for imaging shallow structure I: active source survey, *Chinese Journal of Geophysics*. (in Chinese), **63**(2):532-540, <https://doi.org/10.6038/cjg2020N0184>.
- Uyanik, O. 2019. Estimation of the porosity of clay soils using seismic P- and S-wave Velocities, *Journal of Applied Geophysics*, **170**, 103832, <https://doi.org/10.1016/j.jappgeo.2019.103832>.
- Wang, J., Wu, G., & Chen, X. 2019. Frequency-Bessel transform method for effective imaging of higher-mode Rayleigh dispersion curves from ambient seismic noise data, *Journal of Geophysical Research: Solid Earth*, **124**, 3708-3723, <https://doi.org/10.1029/2018JB016595>.
- Wapenaar, L., der Neut, J., Ruigrok, E., Draganov, D., Hunziker, J., Slob, E., Thorbecke, J. & Snieder, R. 2011. Seismic interferometry by crosscorrelation and by multidimensional deconvolution: a systematic comparison, *Geophysical Journal International*, **185**, 1335-1364, <https://doi.org/10.1111/j.1365-246X.2011.05007.x>.
- West, D. 2020. *Introduction to Graph Theory Second Edition* (in Chinese), China Machine Press.
- Wu, G., Pan, L., Wang, J. & Chen, X. 2020. Shear Velocity Inversion Using Multimodal Dispersion Curves From Ambient Seismic Noise Data of USArray Transportable Array, *Journal of Geophysical Research: Solid Earth*, **125**, e2019JB018213, <https://doi.org/10.1029/2019JB018213>.
- Xi, C., Xia, J., Mi, B., Dai, T., Liu, Y. & Ning, L. 2021. Modified frequency-Bessel transform method for dispersion imaging of Rayleigh waves from ambient seismic noise, *Geophysical Journal International*, **225**, 1271-1280, <https://doi.org/10.1093/gji/ggab008>.
- Xia, J., Miller, R.D. & Park, C.B. 1999. Estimation of near-surface shear-wave velocity by inversion of Rayleigh waves, *Geophysics*, **64**(3), 691-700, <https://doi.org/10.1190/1.1444578>.
- Xia, J., Miller, R.D., Park, C.B., & Tian, G. 2003. Inversion of high frequency surface waves with fundamental and higher modes, *Journal of Applied Geophysics*, **52** (1), 45–57, [https://doi.org/10.1016/S0926-9851\(02\)00239-2](https://doi.org/10.1016/S0926-9851(02)00239-2).
- Yan, Y., Wang, Z., Li, J., Huai, N., Liang, Y., Song, S., Zhang, J., & Zhang, L. 2020. Elastic SH- and Love-wave Full-Waveform Inversion for shallow shear wave velocity with a preconditioned technique, *Journal of Applied Geophysics*, **173**, 103947, <https://doi.org/10.1016/j.jappgeo.2020.103947>.
- Yan, Y., Li, J., Huai, N., Guan, J., & Liu, H. 2022. Two-station analysis of passive surface waves with continuous wavelet transform and plane-wave-based beamforming, *Journal of Applied Geophysics*, **197**, 104526, <https://doi.org/10.1016/j.jappgeo.2021.104526>.
- Yang, B., Xiong, Z., Zhang, D., & Yang, Z. 2019. Rayleigh surface-wave dispersion curve inversion based on adaptive chaos genetic particle swarm optimization algorithm, *Oil Geophysical Prospecting*. (in Chinese), **54**(6), 1217-1227, <https://doi.org/10.13810/j.cnki.issn.1000-7210.2019.06.005>.
- Yilmaz, Ö., Eser, M., & Berilgen, M. 2006. A case study of seismic zonation in municipal areas, *The Leading Edge*, **25** (3), 319–330, <https://doi.org/10.1190/1.2184100>.
- Zhang, K., Li, H., Wang, X. & Wang, K. 2020. Retrieval of shallow S-wave velocity profiles from seismic reflection surveying and traffic-induced noise, *Geophysics*, **85**(6), EN105-EN117, <https://doi.org/10.1190/geo2019-0845.1>.
- Zhang, S., & Chan, L. 2003. Possible effects of misidentified mode number on Rayleigh wave inversion, *Journal of Applied Geophysics*, **53**, 17-29, [https://doi.org/10.1016/S0926-9851\(03\)00014-4](https://doi.org/10.1016/S0926-9851(03)00014-4).

Zhou, J. & Chen, X. 2021. Removal of Crossed Artifacts from Multimodal Dispersion Curves with Modified Frequency–Bessel Method, Bulletin of the Seismological Society of America, <https://doi.org/10.1785/0120210012>.

Zywicki, D., & Rix, G. 2005. Mitigation of the Near-Field Effects for Seismic Surface Wave Velocity Estimation with Cylindrical Beamformers, Journal of Geotechnical and Geoenvironmental Engineering, **131**(8), 970-977, [https://doi.org/10.1061/\(ASCE\)1090-0241\(2005\)131:8\(970\)](https://doi.org/10.1061/(ASCE)1090-0241(2005)131:8(970))

APPENDIX A: Construct the initial model

For the inversion of the measured surface wave dispersion curves, we establish the initial model based on the phase velocities of the fundamental-mode surface wave. We choose the initial model as a homogeneous half-space model or an incremental model with equal layer thickness, and the model is described by the S-wave velocity and thickness of the strata. Supposing that the observed phase velocity vector of the fundamental-mode surface wave is \mathbf{v}_p with the frequency vector \mathbf{f} , then the detection depth d of the initial model is derived from the mean wavelength λ_{mean} of the fundamental-mode surface wave, its expression is:

$$d = \gamma \lambda_{mean} = \gamma \frac{1}{M} \sum_{i=0}^{M-1} \frac{\mathbf{v}_p(f_i)}{f_i}, \quad (\text{A1})$$

where M is the number of the observed values, f_i is the corresponding frequency, γ is an empirical factor ranging from 0.5 to 2.0. If the stratum is divided into l ($l > 1$) layers, and the thickness h_i of the i -th layer is defined as:

$$h_i = \frac{d}{l-1} \quad (i = 0, 1, \dots, l-2). \quad (\text{A2})$$

If the homogeneous half-space model is chosen, the S-wave velocity v_i of the i -th layer is expressed as:

$$v_i = 0.5 \frac{\mathbf{v}_p(\min(\mathbf{f})) + \mathbf{v}_p(\max(\mathbf{f}))}{0.88} \quad (i = 0, 1, \dots, l-1), \quad (\text{A3})$$

where $\mathbf{v}_p(\min(\mathbf{f}))$ represents the phase velocity at the minimum frequency $\min(\mathbf{f})$, and $\mathbf{v}_p(\max(\mathbf{f}))$ is the phase velocity of the maximum frequency. For the incremental model, the expression of v_i is:

$$v_i = v_0 + i\Delta v = \frac{\mathbf{v}_p(\max(\mathbf{f}))}{0.88} + i \frac{\mathbf{v}_p(\min(\mathbf{f})) - \mathbf{v}_p(\max(\mathbf{f}))}{0.88(l-1)} \quad (i = 0, 1, \dots, l-1). \quad (\text{A4})$$

APPENDIX B: Calculation process for the bipartite G

In the section 2.2, we give the calculation steps for solving the maximum matching of a weighted bipartite graph. The Step 2 of the solution process would continue to augment the equal subgraph, and drawing the equal subgraph of the intermediate process in the article is cumbersome. Therefore, we use the matrix to describe the calculation process. The weight matrix \mathbf{W} of bipartite graph G mentioned in section 2.2 is expressed as:

$$\mathbf{W} = \begin{bmatrix} -289 & -9 & -784 \\ -900 & -100 & -225 \end{bmatrix}. \quad (\text{B1})$$

The element $W(i, j)$ at the position (i, j) in the matrix represents the negative distance based on the L^p norm of the i -th measured value with the NESMO and j -th used predicted value. We associate the vertices and labels (a, b) with the rows and columns of the matrix, that is, the vertices set of X and Y are regarded as the rows and columns of the matrix, respectively. Subtracting $W(i, j)$ from $a[i]+b[j]$ to get the corresponding difference matrix. The edges of the equal subgraph correspond to the 0-element in the difference matrix. Initially, the difference matrix is as follows:

$$\begin{array}{ccc} 0 & 0 & 0 \\ -9 \begin{bmatrix} 280 & \underline{0} & 775 \\ 800 & 0 & 125 \end{bmatrix}, \\ -100 \end{array}$$

(B2)

where $[-9, -100]$ and $[0, 0, 0]$ denote the corresponding feasible vertex labels for the sets X and Y, respectively. The edge corresponding to the 0-element marked with an underline in the matrix is the maximum matching O determined in the initial state. O is not a complete match, we need to find a vertex cover set Q according to the Step 2, only if Q covers the vertex Y_2 , Q is not trivial. At this time, R is the empty set, $T = \{Y_2\}$, and the $\varepsilon=125$ is calculated by the formula (12). Then, the feasible vertex labels are modified, and the new difference matrix would be:

$$\begin{matrix} 0 & 125 & 0 \\ -134 & \begin{bmatrix} 155 & 0 & 650 \end{bmatrix} \\ -225 & \begin{bmatrix} 675 & 0 & 0 \end{bmatrix} \end{matrix}$$

(B3)

After this step, the equal subgraph is augmented (a new equal subgraph is generated). And the red 0-elements of the matrix denote the result of the matching relationship.

APPENDIX C: Preconditioned steepest-descent (PSD) algorithm

The mode-orders for the multimodal surface wave observations need to be explicitly specified for the PSD algorithm, and the definition for the misfit function φ derived from the L^p norm is as follows:

$$\varphi(\mathbf{m}) = \frac{1}{N^{1/p}} \left\| \mathbf{v}^o - F(\mathbf{v}^{pre}(\mathbf{m})) \right\|_p, \quad (C1)$$

where \mathbf{m} represents the model occurred in the inversion process, N is the number of observed values, \mathbf{v}^o is the observed dataset for the phase velocities, $\mathbf{v}^{pre}(\mathbf{m})$ denotes the predicted dataset of the multimodal surface waves calculated from the model \mathbf{m} , and the operator F completes the extraction from the $\mathbf{v}^{pre}(\mathbf{m})$ according to the mode-order and frequency information about \mathbf{v}^o .

We refer to the core idea of the Levenberg-Marquardt (L-M) method of achieving the compromise between the steepest descent direction and Gauss-Newton descent direction by introducing the damping factor μ (Aster et al. 2018) to present the PSD, and the particular iteration sequence for it in the paper is defined as:

$$\left\{ \begin{array}{l} \mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{d}_k^N \quad k \geq 0, \\ \mathbf{d}_k^N = (\mathbf{d}_k^0, \mathbf{d}_k^1), \\ \mathbf{d}_k^0 = \frac{\min(\mathbf{V}_{S_k})}{\max(|\mathbf{d}_k(0:l-1)|)} \mathbf{d}_k(0:l-1), \\ \mathbf{d}_k^1 = \frac{\min(\mathbf{h}_k)}{\max(|\mathbf{d}_k(l:2l-2)|)} \mathbf{d}_k(l:2l-2), \\ \mathbf{d}_k = -(\mathbf{H}_k + \mu_k \max(|\mathbf{H}_k|))^{-1} \mathbf{g}_k, \end{array} \right. \quad (C2)$$

where \mathbf{m}_k is the model at the k -th iteration, which consists of the S-wave velocity and layer thickness vector. l is the number of layers, for the k -th iteration, \mathbf{d}_k^N is the amplitude-preconditioned descent direction consisting of the \mathbf{d}_k^0 and \mathbf{d}_k^1 , \mathbf{d}_k^0 and \mathbf{d}_k^1 are the descent direction for the S-wave velocity and layer thickness after the scaled operation according to the corresponding minimum, respectively, \mathbf{g}_k and \mathbf{H}_k are the gradient vector and Hessian matrix of the misfit function, respectively, only the diagonal elements of the \mathbf{H}_k are estimated in the paper. μ_k is the damping factor that ranges from 10^{-6} ~10 in the article, which is determined by the trials at the k -th iteration, and α_k is the inexact percentage step-length (expressed as a percentage of the respective minimum values of the current S-wave velocity and layer thickness vector) given by the Armijo search (Nocedal & Wright 2006). Experimenting with different damping factor μ_k at the k -th iteration would result in different descent directions, that is, giving more options for the model updating for the k -th iteration, which is expected to improve the convergence, stability and sustainability of the inversion.

Accuracy of the partial derivatives of the misfit is key in determining modifications to the model parameters and dramatically affects convergence of the inverse procedure (Xia et al. 1999). Xia et al. (1999) use the Ridder's method of polynomial extrapolation (Press et al. 1992) to calculate the partial derivatives during the inversion. Meanwhile, Xia et al. (1999) also pointed out that the practical way to calculate the partial derivatives of the misfit is by evaluating finite-difference values because the misfit is an implicit function about the model parameters. Therefore, we simply calculate the partial derivatives of the misfit through the numerical difference, then, the i -th element $\mathbf{g}_k(i)$ of the gradient vector at the k -th iteration is approximated by the central difference format:

$$\mathbf{g}_k(i) = \frac{\varphi(\mathbf{m}_k + \varepsilon_i \mathbf{e}_i) - \varphi(\mathbf{m}_k - \varepsilon_i \mathbf{e}_i)}{2\varepsilon_i} \quad (i = 0, 1, \dots, 2l-2), \quad (\text{C3})$$

where the ε_i is the perturbation parameter at the i -th dimension about the current model \mathbf{m}_k , \mathbf{e}_i is the base vector for the corresponding dimension, and the expressions for them are:

$$\begin{cases} \varepsilon_i = 0.005\mathbf{m}_k(i), \\ \mathbf{e}_i = (0, \dots, 1, \dots, 0)^T. \end{cases} \quad (\text{C4})$$

In the above formula, the element 1 is placed at the i -th position of the \mathbf{e}_i , and the symbol T represents the transpose operation. Similarly, the diagonal element $\mathbf{H}_k(i, i)$ of the Hessian matrix at the k -th iteration is estimated by:

$$\mathbf{H}_k(i, i) = \frac{\varphi(\mathbf{m}_k + \varepsilon_i \mathbf{e}_i) + \varphi(\mathbf{m}_k - \varepsilon_i \mathbf{e}_i) - 2\varphi(\mathbf{m}_k)}{\varepsilon_i^2} \quad (i = 0, 1, \dots, 2l-2). \quad (\text{C5})$$

To sum up, the detailed steps for the implementation of the multimodal surface wave dispersion inversion based on the PSD is as follows:

Algorithm 3: Preconditioned steepest-descent (PSD) algorithm

Step 1: Let $k = 0$, given the initial model \mathbf{m}_k and the maximum number of the iterations $maxIter$, calculating the misfit value φ_k of the \mathbf{m}_k .

Step 2: If $k > maxIter$, the iteration is terminated and the inverted result is returned.

Step 3: Calculating the gradient vector \mathbf{g}_k and the Hessian matrix \mathbf{H}_k of the misfit function according to the formula (C3)~(C5).

a) Outer loop about the damping factor μ_k : Given the sequence of the μ_k for the trial tests, the sequence for it is set as $(10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1.1, 10)^T$ in the paper. Then, the amplitude-preconditioned descent direction \mathbf{d}_k^N is obtained for each trial μ_k .

b) Inner loop about the step-length α_k : To perturb the current model \mathbf{m}_k with the inexact step-length and the trial descent direction \mathbf{d}_k^N , the sequence of the step-length is given as

$(0.0125/4, 0.0125/2, 0.0125, 0.025, 0.05)^T$ in the paper, there would be 5 trial models generated by the given step-length sequence with each trial θ_k , then calculating the misfit values of these models and saving them.

Step 4: To find the smallest misfit value φ_{min} of these trial models occurred in the Step 3 and its corresponding model \mathbf{m}_{min} . If $\varphi_{min} < \varphi_k$, let $k=k+1$, $\varphi_k=\varphi_{min}$, $\mathbf{m}_k=\mathbf{m}_{min}$, then go back to the Step 2 and 3; otherwise, the iteration is terminated and return the inverted result \mathbf{m}_k .

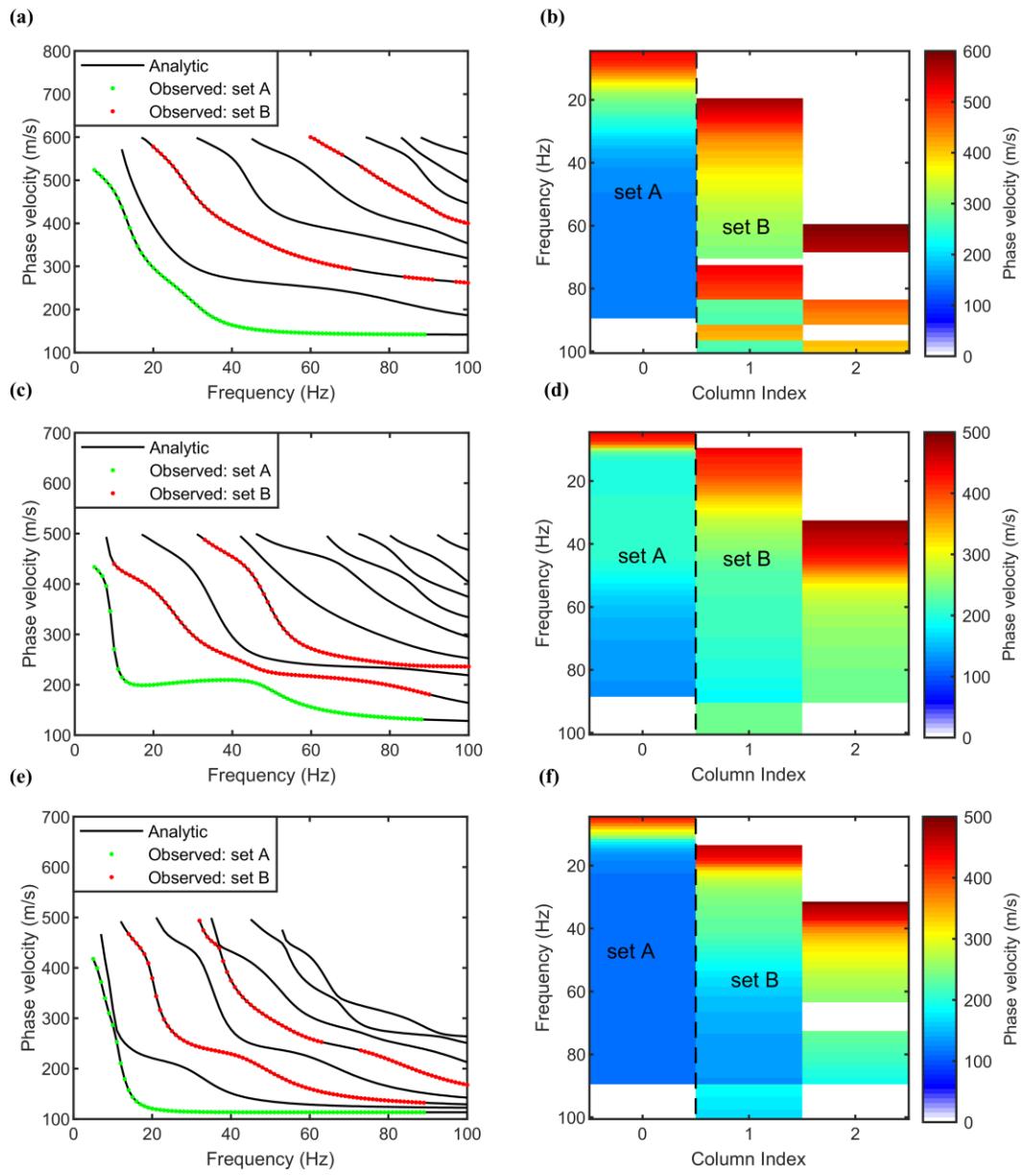


Figure 1. The observed phase velocities and data structures for the synthetic tests. (a) and (b) are prepared for the synthetic reconstruction test of the four-layer incremental model (Model A). (c) and (d) are used for the four-layer model with a low-velocity weak interlayer (Mode B) reconstruction. (e) and (f) are prepared for the reconstruction of the four-layer model with a high-velocity hard interlayer (Model C).

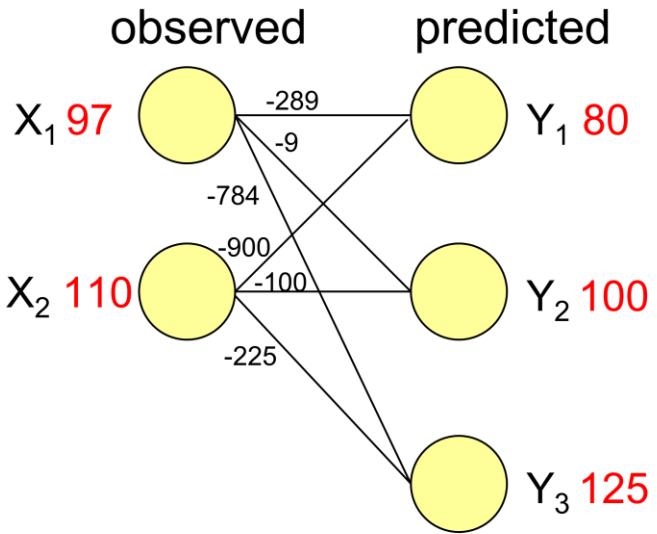


Figure 2. A schematic bipartite graph G , and the yellow solid circles denote the vertices. The vertices of the observed values with the NESMO and the used predicted values at each frequency are represented by the sets X and Y , the red numbers denote the corresponding phase velocities, connecting the vertices of the observed values to all the used predicted values at a certain frequency, and the weights of the edges are calculated by the negative distance based on the L^p (p is set as 2) norm between the vertices of X and Y . The numbers on each edge represent the weight.

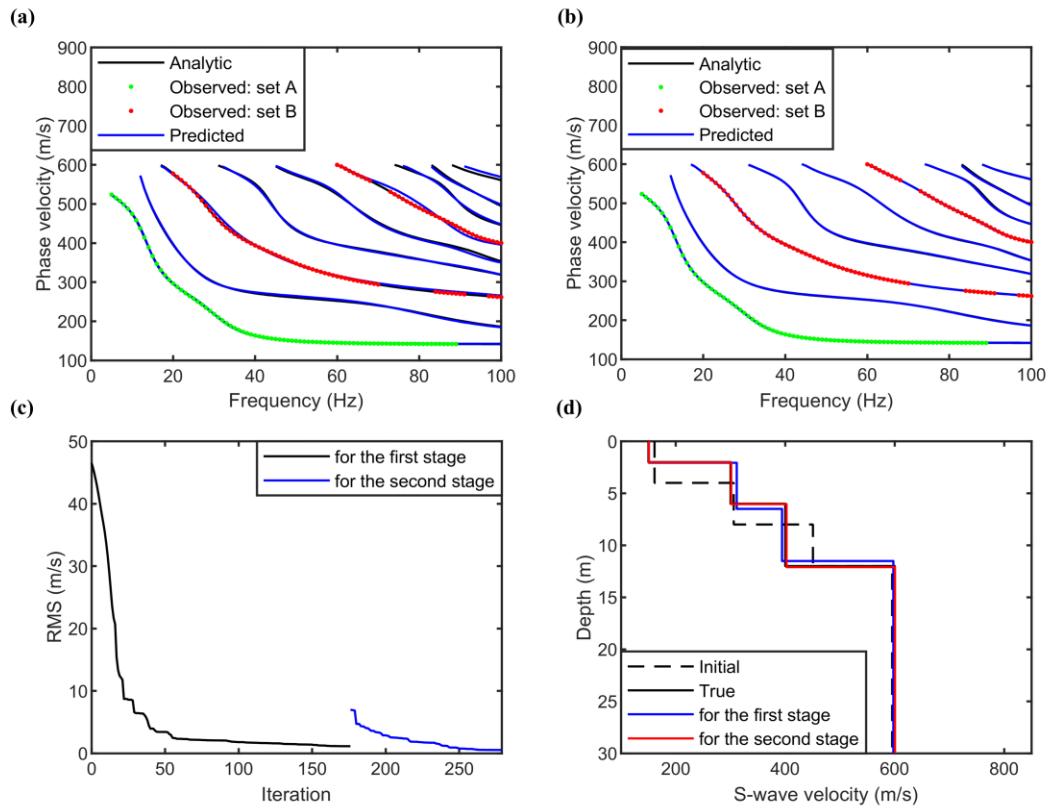


Figure 3. The evaluation curves and inversion result for the synthetic test of the Model A with error-free data. The comparison (a) of the analytic, observed and predicted phase velocities calculated from the inversion result of the first stage, and (b) is the comparison of those curves of the second stage. The descent curves (c) of the misfit values against the iteration for the first and second stage. (d) The initial, true, and the reconstructed 1D S-wave velocity structures of the first and second stage.

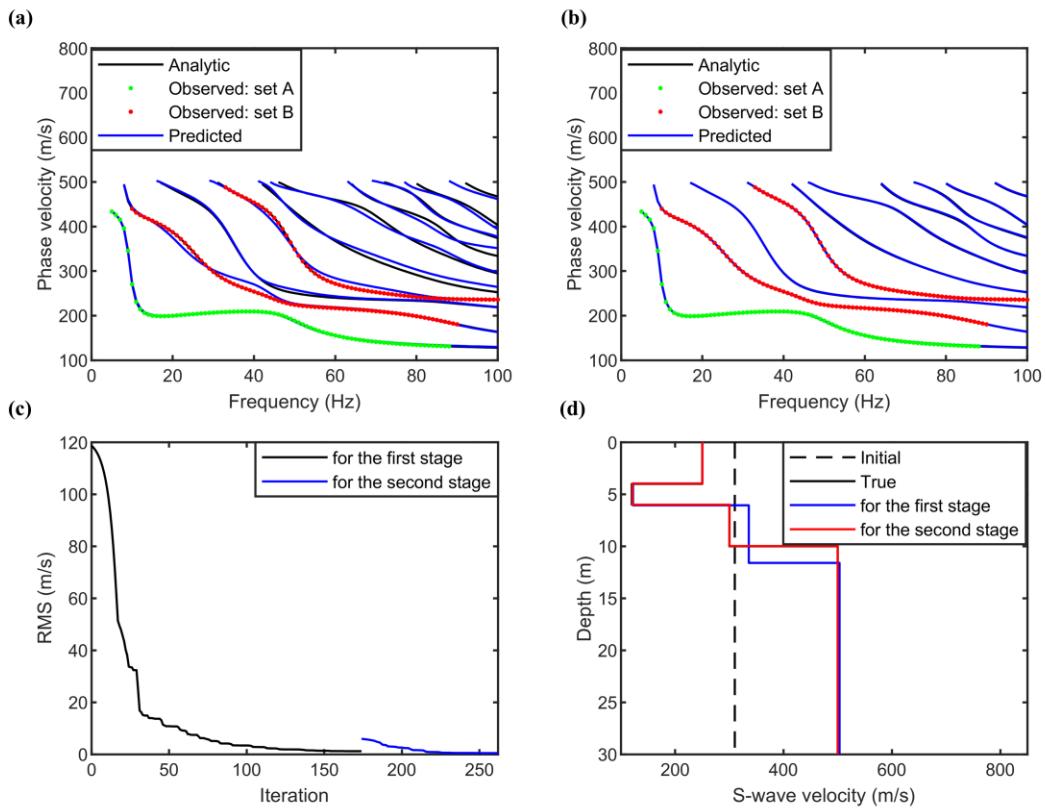


Figure 4. The evaluation curves and inversion result for the synthetic test of the Model B with error-free data. The comparison (a) of the analytic, observed and predicted phase velocities calculated from the inversion result at the first stage, and (b) is the comparison of those curves at the second stage. The descent curves (c) for the misfit values of the first and second stage. (d) The comparison for the initial model, true model, and inverted results of the first and second stage.

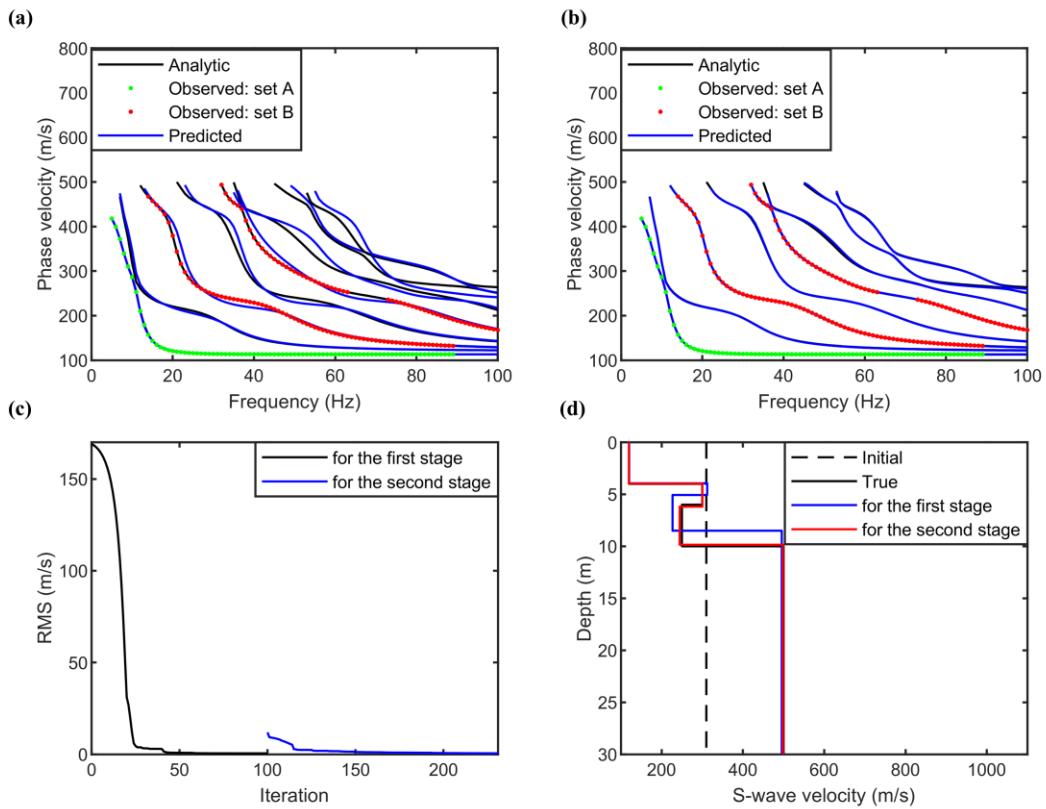


Figure 5. The evaluation curves and inversion result for the synthetic test of the Model C with error-free data. The comparison (a) of the analytic, observed and predicted phase velocities calculated from the inversion result of the first stage, and (b) is the comparison of those curves of the second stage. The descent curves (c) for the misfit values of the first and second stage. (d) The comparison for the initial model, true model, and the inverted result of the first and second stage.

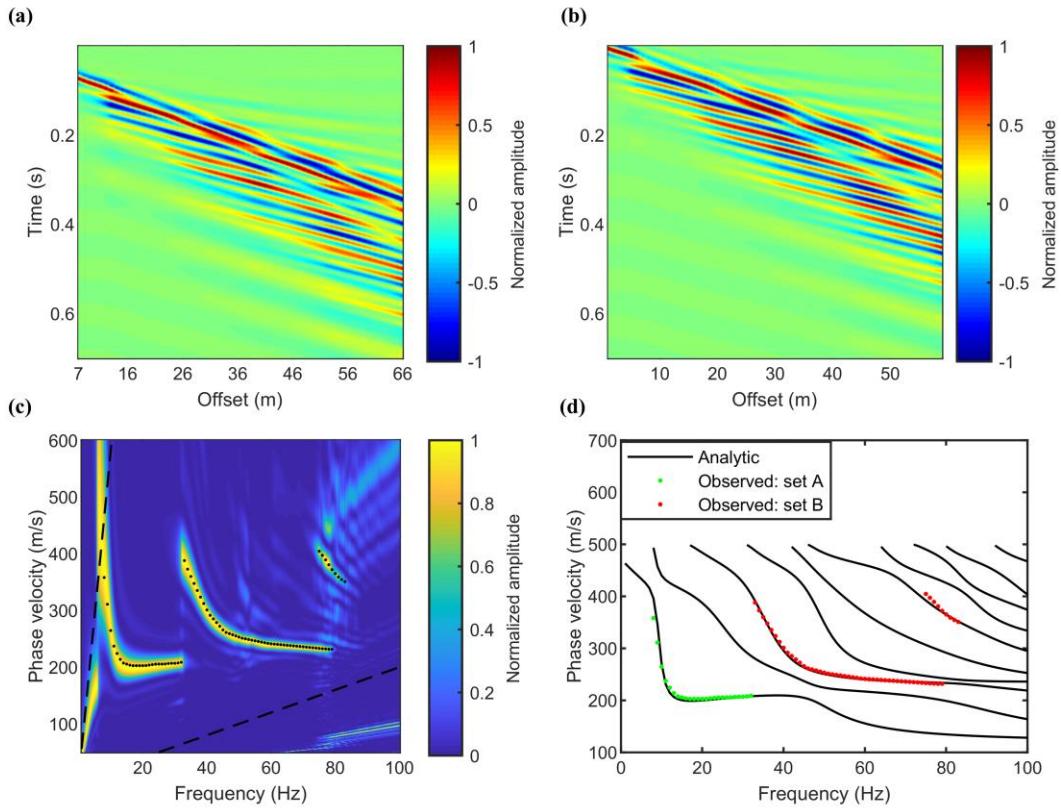


Figure 6. The synthetic Z-component records (a) for the Model B described in Table 2, each trace is normalized with respect to its maximum absolute amplitude. The cross-correlation function (CCF) gather (b) is generated by the cross-correlation with the first trace as the reference, and each trace is also normalized with respect to its maximum absolute amplitude. The energy map (c) in f - v domain is extracted from the CCF gather (b) by the F-J transform, the two black dashed lines restrict a reliable zone between the Nyquist wavenumber ($1/(2*dx)$) (dx is the receiver-interval) and the maximum wave length (the length of the array), and the black dotted lines denote the observed phase velocities with 1 Hz interval. The analytic and observed phase velocities (d) for the Model B, the phase velocities of the set A belong to the fundamental-mode surface wave, and that of the set B are the NESMO.

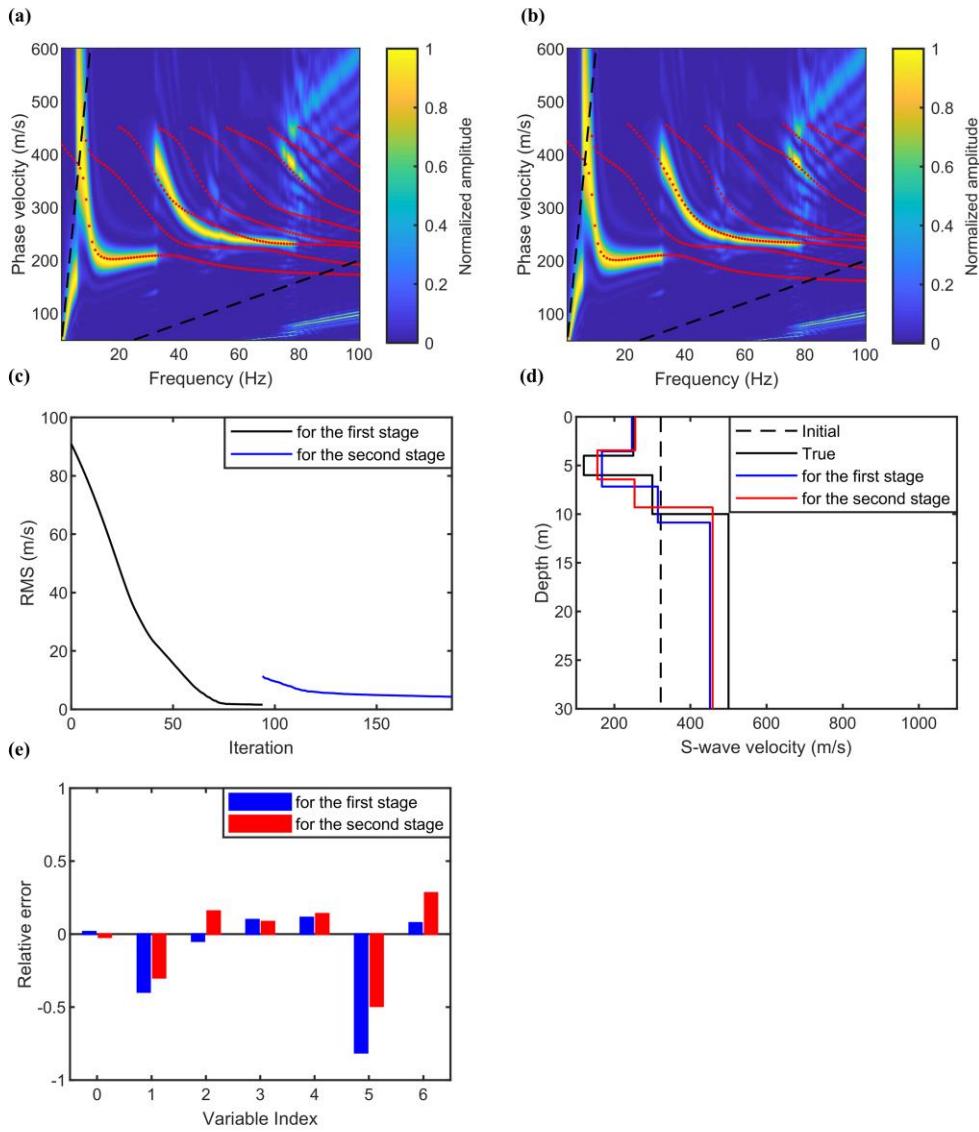


Figure 7. (a) The dispersive spectrum extracted from the CCF gather and the red dotted lines represent the predicted phase velocity values calculated from the inversion result of the first stage. (b) The red dotted lines also denote the predicted phase velocities of the inversion result of the second stage. (c) The descent curves of the misfit of the two stages. (d) The comparison for the initial, true, and inverted result of the first stage and second stage. (e) The distribution of the relative errors for the parameters in all dimensions between the inverted result and true model, and the variable index 0~3 belongs to the S-wave velocity parameter, the variable index 4~6 is for the layer thickness parameter.

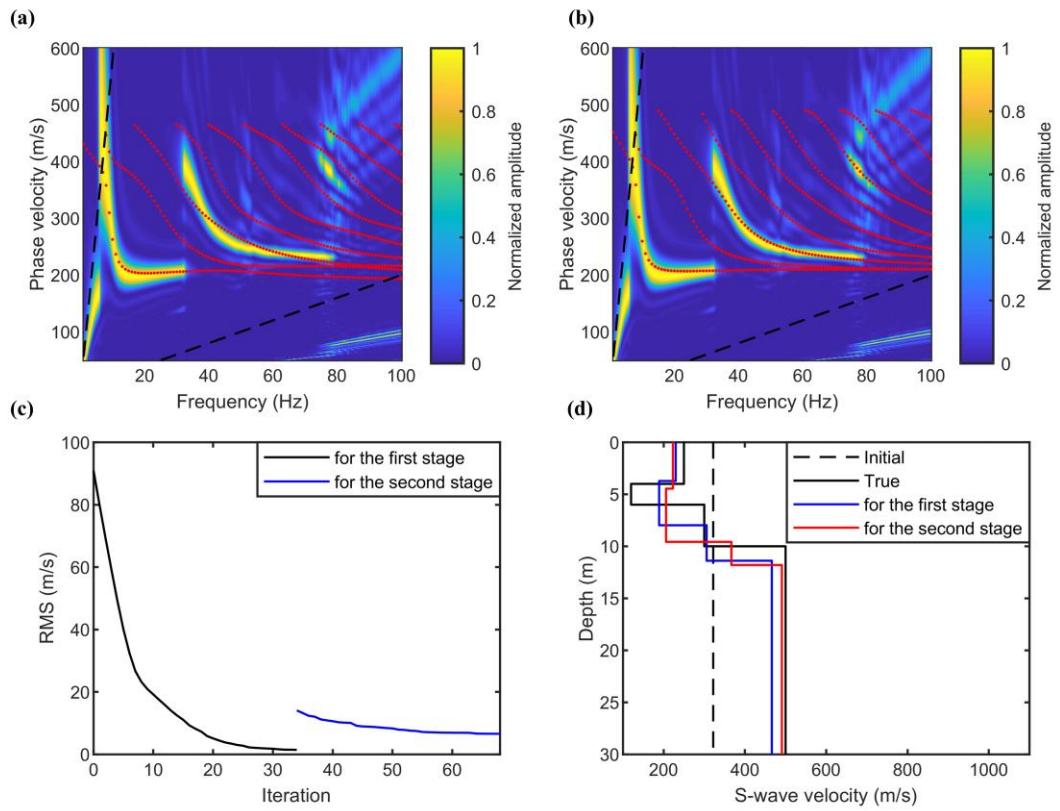


Figure 8. The dispersive spectrum (a) extracted from the CCF gather by the F-J transform, and the red dotted lines represent the predicted multimodal surface wave dispersion curves calculated from the inverted result at the first stage. The red dotted lines on (b) are also the predicted multimodal surface wave dispersion curves, which is calculated from the inverted result at the second stage. (c) The descent curves of the misfit values against the iteration for the first and second stage. (d) The comparison for the initial, true, and inverted result of the first stage and second stage.

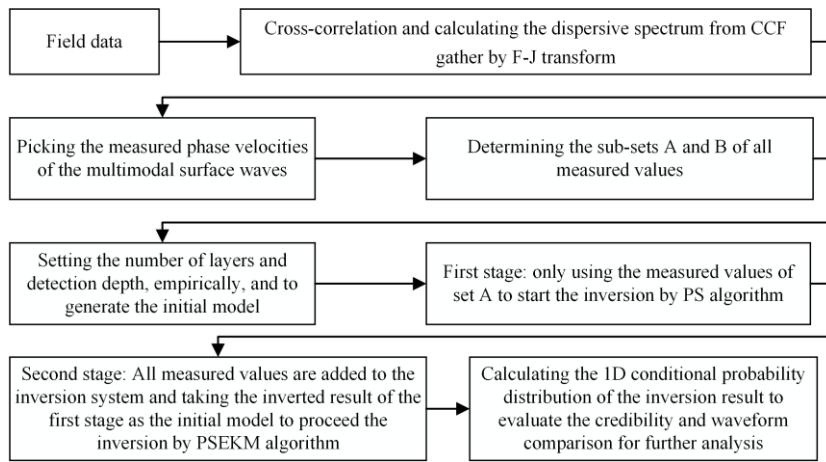


Figure 9. The analysis workflow for the field surface wave data.

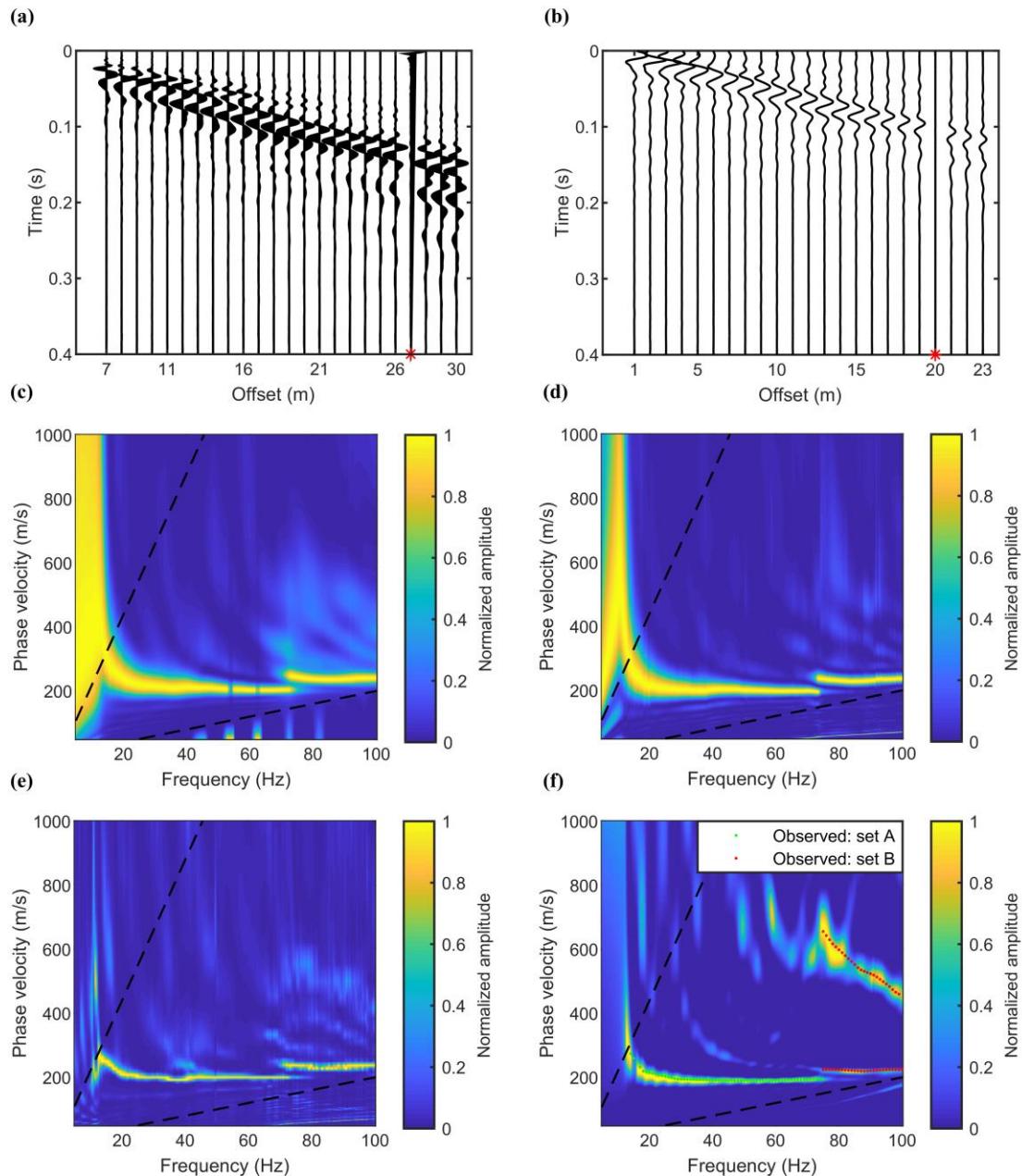


Figure 10. (a) The normalized seismograms of the raw Rayleigh-wave records of roadbed 1, and each channel is normalized with respect to its maximum absolute amplitude. The first trace of (a) is set as the reference to get the cross-correlation function (CCF) gather (b). The red stars on (a) and (b) mark the horizontal location for the borehole operation. The dispersive spectrums generated by the 2D Fourier transform (c), phase-shift (d), HRLRT (e) and the F-J transformation (f). Here, the real part of the F-J spectrum is shown. The two black dashed lines in (c), (d), (e) and (f) restrict a reliable area for picking the phase velocities with the Nyquist wavenumber at $1/(2 \cdot dx)$ and maximum wavelength (equals to the array length). The observed phase velocities

have been picked from (f) with 1 Hz interval, the set A stores the fundamental-mode surface wave observations, and that in the set B are all NESMO.

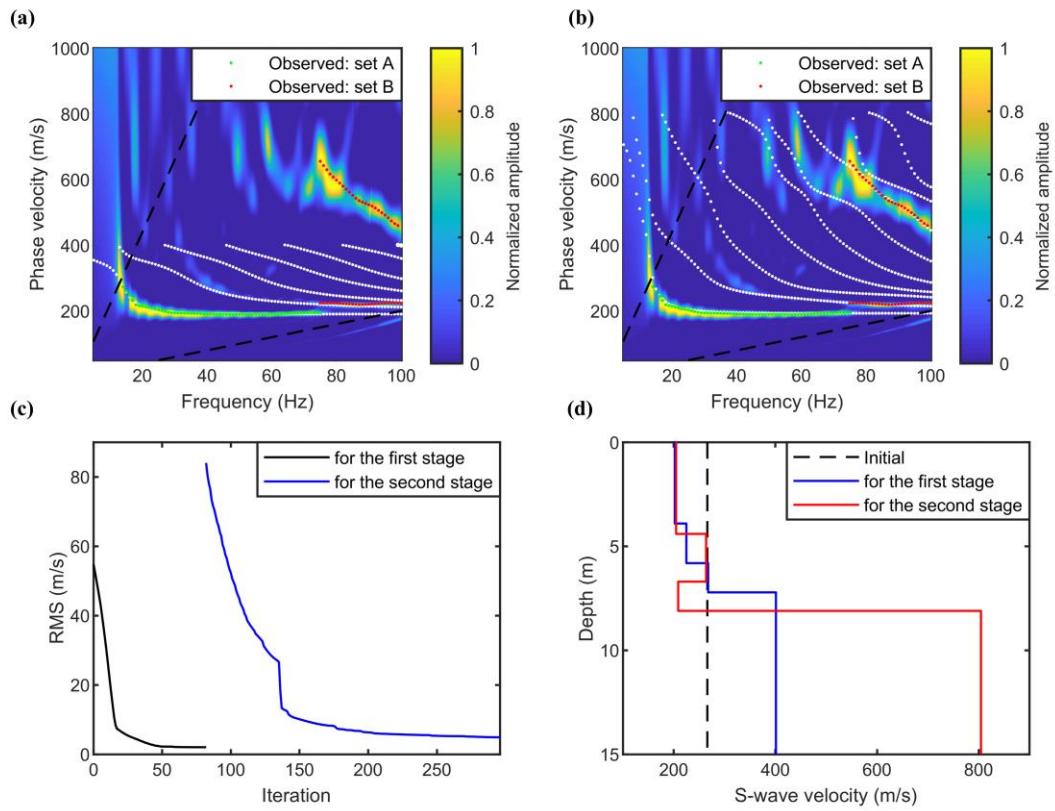


Figure 11. The dispersive spectrum (a) extracted from the CCF gather of roadbed 1 by the F-J transform, and the white dotted lines represent the predicted multimodal surface wave dispersion curves calculated from the inverted result at the first stage. The white dotted lines on (b) are also the predicted multimodal surface wave dispersion curves, which is calculated from the inverted result at the second stage. The two black dashed lines in (a) and (b) define a reliable area for picking the measured values of the phase velocity with the Nyquist wavenumber at $1/(2 \cdot dx)$ and maximum wavelength (equals to the array length). (c) The descent curves of the misfit values against the iteration for the first and second stage. (d) The comparison for the initial, and inverted result of the first stage and second stage.

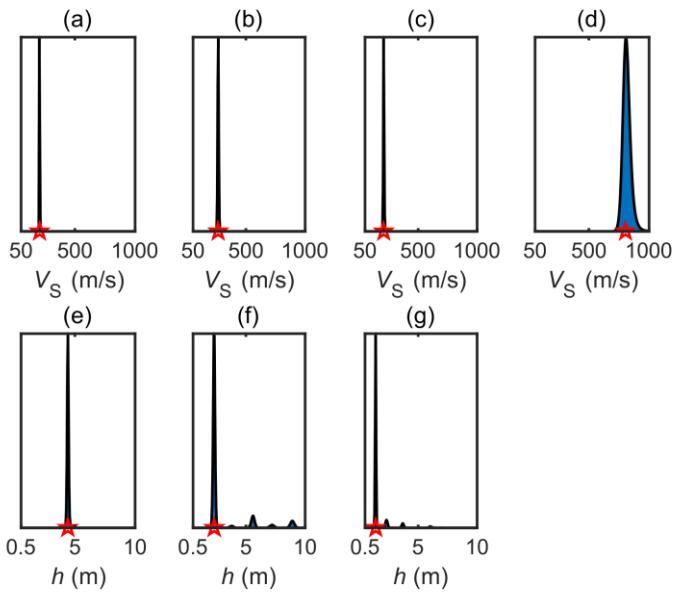


Figure 12. The 1D conditional probability density distribution (CPDD) of the inversion parameters for the second stage at each dimension for roadbed 1. (a)~(d) describe the 1D CPDD for the S-wave velocity of each layer, and (e)~(g) represent that for the formation thickness.

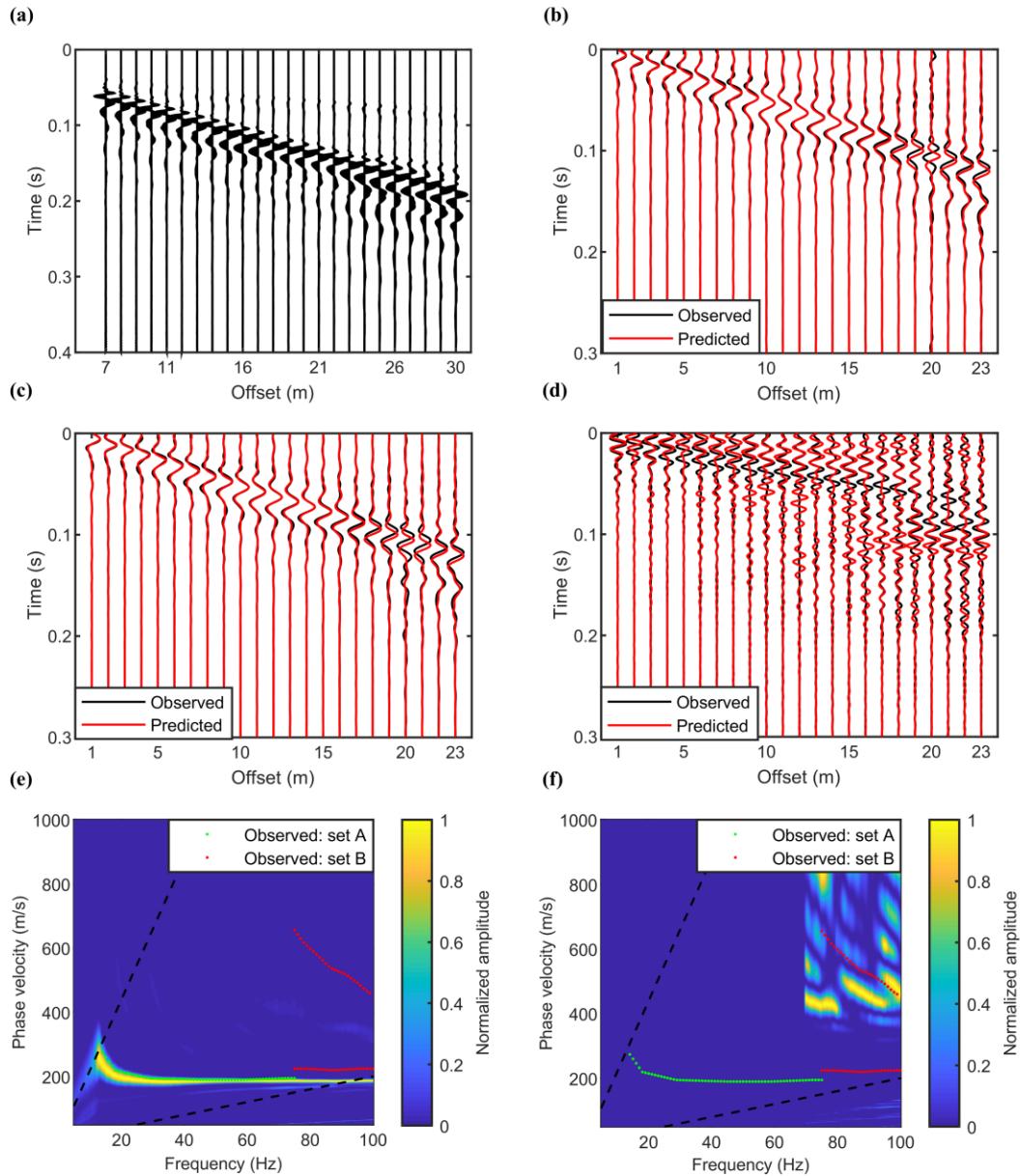


Figure 13. (a) The normalized seismograms of the synthetic Rayleigh-wave records of the inversion result at the second stage for the roadbed 1, and each channel is normalized with respect to its maximum absolute amplitude. (b) The first trace of (a) is set as the reference to get the predicted cross-correlation function (CCF) gather, the black one is the observed CCF waveform, and the red one denotes the predicted CCF. (c) The separated parts of the observed and predicted CCFs with phase velocity less than 400 m/s. (d) The separated parts of the observed and predicted CCFs with phase velocity greater than 400 m/s, then the 65~110 Hz bandpass filter is operated on these CCFs. Each CCF waveform of (b), (c) and (d) are normalized to 0.5 with

respect to corresponding maximum absolute amplitude for clarity. (e) The dispersion image of the predicted CCFs of (c). (f) The dispersion image of the predicted CCFs of (d), and only the dispersion energy of 70~100 Hz is normalized for each frequency. The two black dashed lines in (e) and (f) restrict a reliable area for picking the phase velocities with the Nyquist wavenumber at $1/(2*dx)$ and maximum wavelength (equals to the array length).

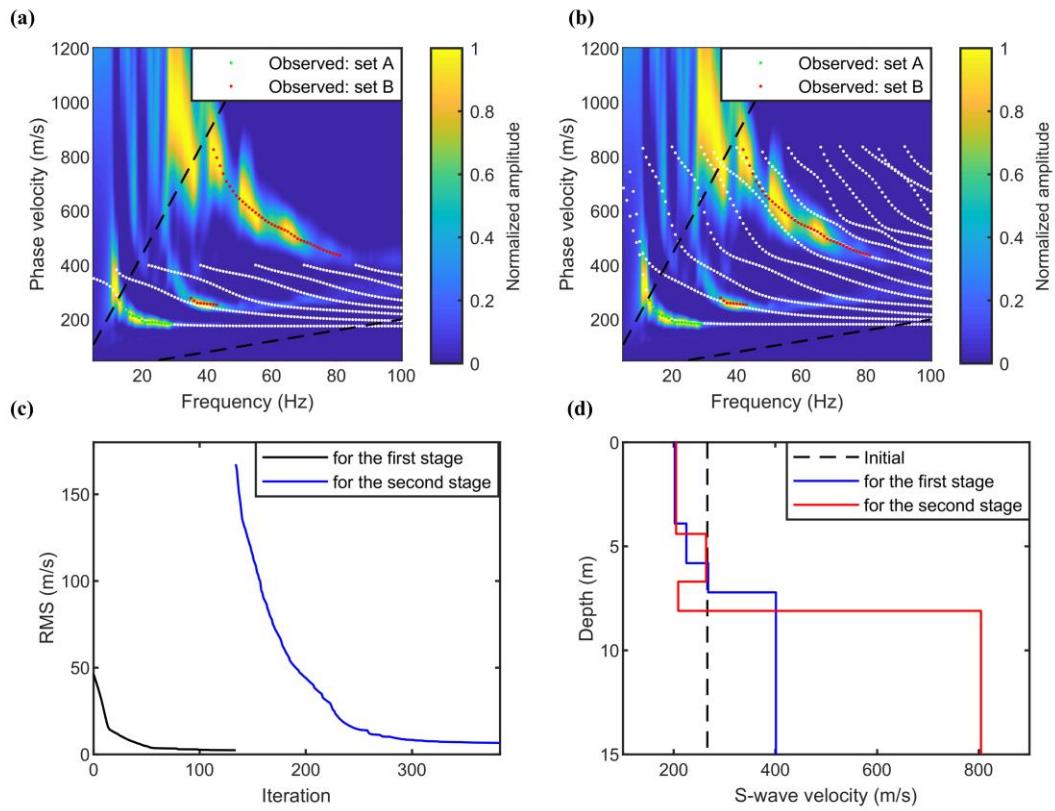


Figure 14. The dispersive spectrums (a) and (b) of the roadbed 2 extracted by the F-J transform, and the white dotted lines on (a) and (b) represent the predicted multimodal surface wave dispersion curves calculated from the inversion result at the first and second stage, respectively. The two black dashed lines in (a) and (b) define a reliable area for picking the phase velocities with the Nyquist wavenumber at $1/(2 \cdot dx)$ and maximum wavelength (equals to the array length). (c) The descent curves for the misfit function values at the first and second stage. (d) The comparison between the initial and inverted results of the first and second stage.

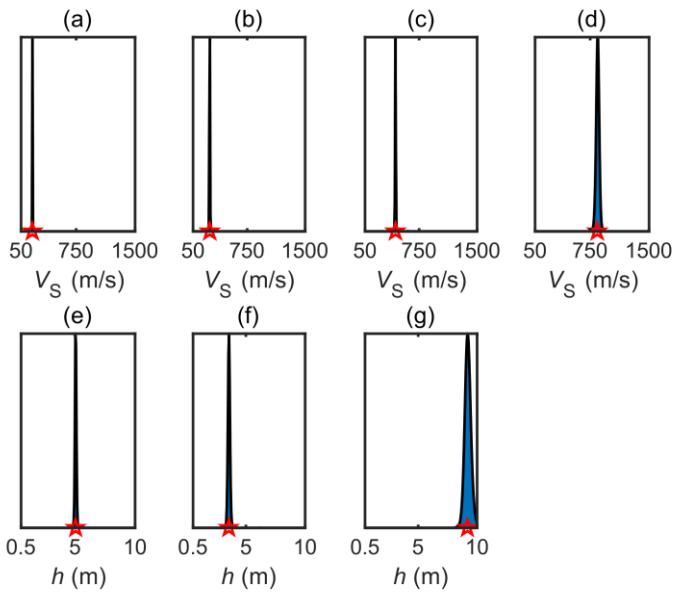


Figure 15. The 1D conditional probability density distribution (CPDD) of the inversion parameters for the second stage at each dimension for the roadbed 2. (a)~(d) describe the 1D CPDD for the S-wave velocity of each layer, and (e)~(g) represent that for the formation thickness parameter.

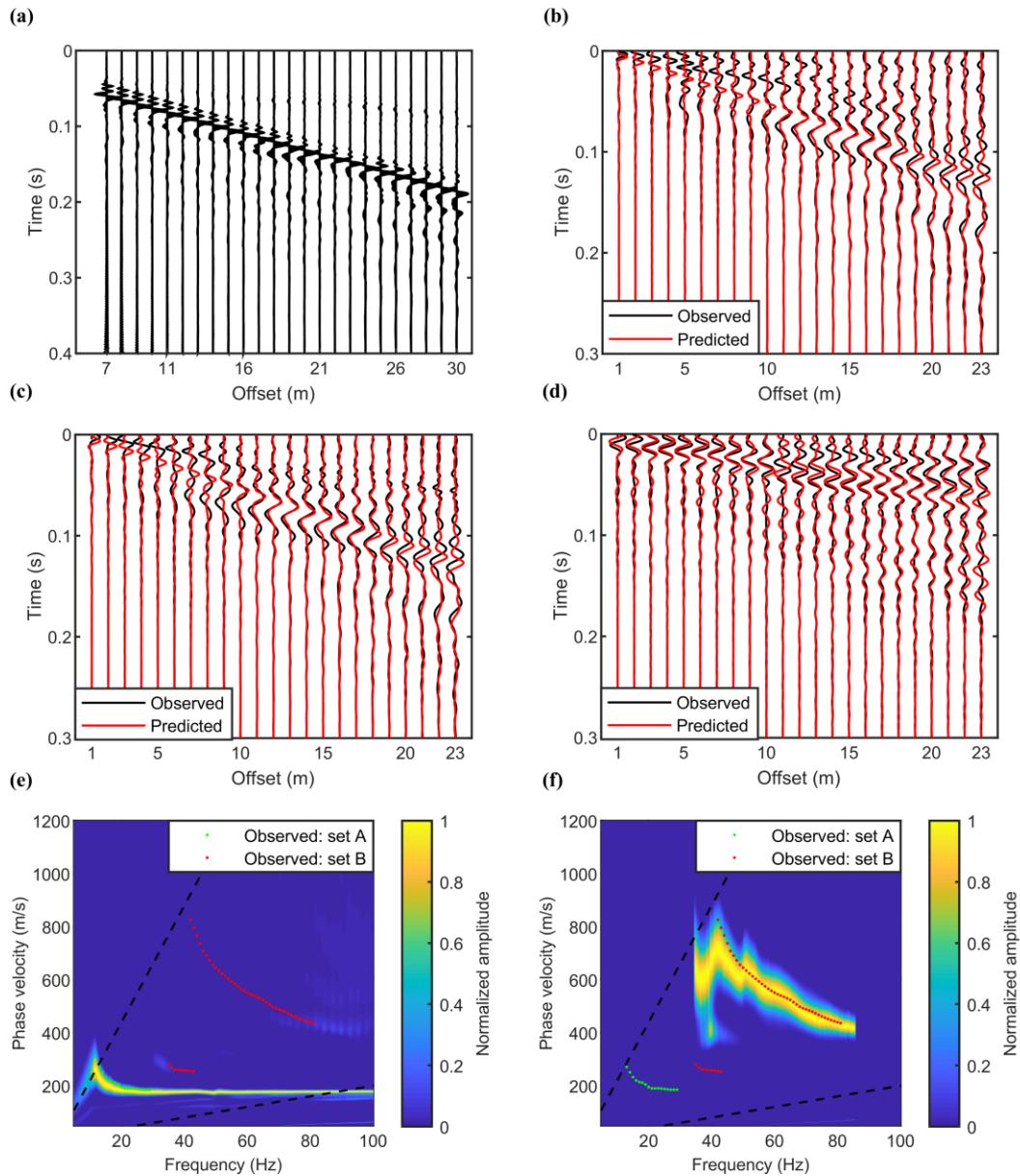


Figure 16. (a) The normalized seismograms of the synthetic Rayleigh-wave records of the inversion result at the second stage for roadbed 2, and each channel is normalized with respect to its maximum absolute amplitude. (b) The first trace of (a) is set as the reference to get the predicted cross-correlation function (CCF) gather, the black and red ones denote the observed and predicted CCFs, respectively. (c) The separated parts of the observed and predicted CCFs with phase velocity less than 400 m/s. (d) The separated parts of the observed and predicted CCFs with phase velocity greater than 400 m/s, then the 35–85 Hz band-pass filter is operated on these CCFs. Each CCF waveform of (b), (c) and (d) are normalized to 0.5 with respect to corresponding maximum absolute amplitude for clarity. (e) The dispersion image of the predicted CCFs of (c).

(f) The dispersion image of the predicted CCFs of (d), and only the dispersion energy of 35~85 Hz is normalized for each frequency. The observed phase velocities for the roadbed 2 are plotted on (e) and (f) for comparison. The two black dashed lines in (e) and (f) restrict a reliable area for picking the phase velocities with the Nyquist wavenumber at $1/(2*dx)$ and maximum wavelength (equals to the array length).

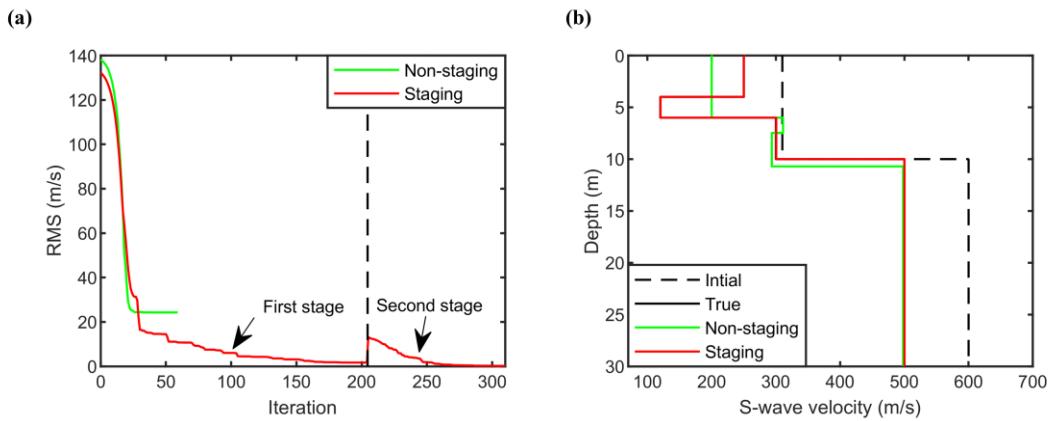


Figure 17. The descent curves (a) of the misfit values with staging and non-staging strategy. The descent curves of the misfit values for the first and second stage are separated by the black dashed line, as marked by the black arrows. (b) The comparison for the initial, true, and the reconstructed S-wave velocity structure obtained by the staging and non-staging strategy.

Table 1 Model A: Parameters of the four-layer model with the incremental velocity structure

Layer number	V_S (m/s)	V_P (m/s)	ρ (kg/m ³)	h (m)	Boundary constraints
1	150	367.5	2000	2	V_S : 100~300 m/s h : 0.5~8 m
2	300	735	2000	4	V_S : 100~800 m/s h : 0.5~8 m
3	400	980	2000	6	V_S : 100~800 m/s h : 0.5~8 m
4	600	1470	2000	∞	V_S : 100~800 m/s

Table 2 Model B: Parameters of the four-layer model with a low-velocity soft interlayer

Layer number	V_S (m/s)	V_P (m/s)	ρ (kg/m ³)	h (m)	Boundary constraints
1	250	612.5	2000	4	V_S : 200~500 m/s h : 0.5~6 m
2	120	294	2000	2	V_S : 100~600 m/s h : 0.5~6 m
3	300	735	2000	4	V_S : 50~800 m/s h : 0.5~6 m
4	500	1225	2000	∞	V_S : 50~800 m/s

Table 3 Model C: Parameters of the four-layer model with a high-velocity hard interlayer

Layer number	V_S (m/s)	V_P (m/s)	ρ (kg/m ³)	h (m)	Boundary constraints
1	120	294	2000	4	V_S : 100~400 m/s h : 0.5~6 m
2	300	735	2000	2	V_S : 200~500 m/s h : 0.5~6 m
3	250	612.5	2000	4	V_S : 50~800 m/s h : 0.5~6 m
4	500	1225	2000	∞	V_S : 50~800 m/s

Table 4 The detailed description of the borehole data and S-wave velocity range for each formation of the roadbed 1

Serial No.	From (m)	To (m)	Formation and description	S-wave velocity (m/s)
1	0.00	3.90	Silty clay, brownish-yellow, uniform in color, of which 0.00-0.50 m is planting soil.	120~400 m/s
2	3.90	6.90	Silty clay, brown, with powder core.	120~400 m/s
3	6.90	9.10	Loose pebble soil, yellowish-brown, composed of quartz, mainly sandstone, filled with fine sand.	200~230 m/s
4	9.10	9.70	Amphibolite, yellowish-brown, weakly weathered, unrecognizable lithofacies structure, rock and soil sandwiched with gravel.	500~1000 m/s
5	9.70	20.50	Amphibolite, yellowish-brown, weakly weathered, dominated by acid plagioclase, followed by quartz.	500~1000 m/s