doi: 10.11720/wtyht.2026.1414
赵云龙, 孙成禹, 黄宏宇. 一种基于能量密度空间聚类算法的面波频散曲线自动拾取方法[J]. 物探与化探, 2026, 50(3): 507-515. https://doi.org/10.11720/wtyht.2026.1414
Zhao Y L, Sun C Y, Huang H Y. A method for automatically picking surface wave dispersion curves based on the energy density-based spatial clustering of applications with noise algorithm[J]. Geophysical and Geochemical Exploration, 2026, 50(3): 507-515. https://doi.org/10.11720/wtyht.2026.1414

一种基于能量密度空间聚类算法的面波频散曲线自动拾取方法

赵云龙, 孙成禹, 黄宏宇
(中国石油大学(华东)地球科学与技术学院, 山东 青岛 266580)

摘 要: 面波分析方法目前已广泛应用于许多不同领域, 用于表征近地表的横波速度结构。面波频散信息的提取是面波分析方法的重要环节之一, 传统的频散曲线拾取方法主要是通过手动拾取, 或人为划定频散能量区域半自动拾取。对于油气勘探中地震数据量大的特点, 人工处理方法无法达到实际生产要求。为此, 本文采用一种基于无监督机器学习方法来自动拾取多阶频散曲线。首先, 利用能量阈值滤波去除一部分频散谱中的背景噪声点。然后通过基于能量密度的噪声应用空间聚类(E-DBSCAN)算法将其区分为不同的面波模式。最后搜索不同面波模式区域中的峰值, 并通过卡尔曼滤波进行平滑处理来消除噪声的影响, 获得多模式频散曲线。该方法考虑了频散能量值, 具有较高的可靠性。数值实验表明, 自动拾取的频散曲线能与理论频散曲线准确地匹配。同时, 方法对比及实际数据测试进一步证实了方法的可靠性。

关键词: 面波; 频散曲线; 聚类算法; 自动拾取
中图分类号: P631.4; P31 **文献标识码:** A **文章编号:** 1000-8918(2026)03-0507-09

0 引言

面波分析方法目前已成为浅地表探测的常用技术, 通过反演频散曲线获得地下横波速度结构。该方法通过频散成像技术(例如相移法、HRLRT 法和 fv-MUSIC 法等)[1-3] 将地震数据从时间—空间域转换到频率—速度域, 利用生成的频散谱识别多模式频散能量并拾取频散曲线, 为构建横波速度模型提供关键依据[4-6]。与基阶频散曲线反演相比, 多模式频散曲线反演提供了更高的分辨率和更大的反演深度, 因此准确提取多阶频散曲线是面波分析方法中的关键环节。

传统频散曲线拾取方法依赖于频散谱中局部能量极大值的视觉识别, 需人工追踪各频率的能量峰值以拾取频散曲线, 或者通过手动划定频散能量区域实现半自动提取[7]。近年来, 随着节点地震仪、光纤传感等先进地震设备的应用, 地震数据量激增, 大量的交互工作不仅耗费大量人力和时间, 还容易引入人为错误。特别是对于大范围的石油勘探, 需要提取大量的频散曲线用于反演, 手动提取频散曲线繁琐又不切实际。因此为了提高数据处理和反演的效率和准确性, 研究多阶频散曲线的自动拾取方法至关重要。

机器学习方法已经被广泛应用于地球物理领域[8-11], 基于无监督学习的频散曲线自动拾取方法也引起了学者关注。Ortega 等[12] 采用了一种基于贝叶斯框架的策略, 以优化频散曲线的获取过程。Xie 等[13] 提出搜索局部最大值的方法拾取基阶频散曲线。Taipodia 等[14] 提出通过对频散谱施加高能量阈值自动拾取基阶频散曲线, 但结果易受噪声干扰。Rovetta 等[15] 提出了一种基于密度的无监督空间聚类算法, 用于自动拾取基阶频散曲线, 并成功将该算法应用于实际数据中。Wang 等[16] 结合高斯混合模

收稿日期: 2025-11-28; **修回日期:** 2026-03-15
基金项目: 国家自然科学基金项目"强吸收介质中面波和 P-导波频散—衰减特征及近地表参数一体化反演"(42574157)
第一作者: 赵云龙(2001-), 男, 硕士研究生, 主要从事面波成像方面的研究工作。Email: zhao.yun.long@outlook.com

型算法与基于噪声应用的密度空间聚类算法，实现了多阶频散曲线的自动提取。Liu 等[17] 提出了自动频散曲线自适应拾取策略(adaptive automatic picking strategy of dispersion curve, AAPDC)，通过基于频散谱局部能量最大值的位置自动搜索频散曲线。Yao 等[18] 采用了 K 均值聚类算法、主成分分析和局部离群因子检测等无监督学习算法组合了频散曲线自动拾取方法。康娜等[19] 提出了一种基于区域生长算法的基阶频散曲线自动拾取方法。Kuang 等[20] 提出了一种两步聚类方法，实现了从多分量数据中拾取多阶频散曲线。陈杰等[21] 采用一种多重滤波法拾取面波多阶频散曲线。宋欣悦等[22] 在多道面波分析法中引入基于曼哈顿距离的 K-Means 算法来拾取频散曲线，提高频散曲线拾取的准确性。

深度学习是机器学习的一个分支，它通过训练神经网络模型来实现频散曲线的拾取。Alyousuf 等[23] 使用具有多个隐藏层的神经网络实现了基阶频散曲线的拾取。Dai 等[24] 开发了一种深度学习模型，通过将频散曲线提取视为实例分割任务，能够快速拾取基阶频散曲线。伍锡军[25] 利用 U-net 神经网络，在面波地震记录中实现了基阶频散曲线的准确拾取。Zhang 等[26] 利用 U-net 网络并结合迁移学习，实现了多阶频散曲线提取与模式识别。然而，神经网络的训练需要大量手动拾取的频散曲线作为数据集，如果训练样本不够丰富，模型的泛化能力可能会受到限制，导致拾取的频散曲线不准确。

本文提出了一种基于能量密度空间聚类(E-DBSCAN)算法的多阶面波频散曲线自动拾取方法。首先，通过相移法将地震数据转换为频散谱，使用能量阈值滤波去除一部分频散谱中的背景噪声点。然后，应用 E-DBSCAN 算法来区分不同模式的频散能量。通过搜索相应模式区域中的局部峰值，可以获得频散曲线。最后，使用卡尔曼滤波对初步提取的频散曲线进行平滑处理，消除噪声的干扰。理论模型测试验证了该方法的有效性和准确性。

1 方法原理

1.1 E-DBSCAN 算法

本文提出使用基于能量密度的噪声应用空间聚类(E-DBSCAN)算法自动拾取频散谱的频散曲线。该算法主要包含几个核心概念：

1) 邻域：对于数据集中的任意点 p，以其为中心、指定半径 ε 构成的圆形区域称为该点的邻域。点 p 的 ε-邻域(即所有与 p 距离不超过 ε 的点)可表示为：

$$N_\varepsilon(p) = \{q \in D | d(p,q) \leqslant \varepsilon\}, \tag{1}$$

式中：D 代表点集；$d(p,q)$表示点 p 与点 q 之间的欧式距离。

2) 核心点：如果在某个点的 ε-邻域内点的能量总和至少达到 *MinEnergy*，则该点可被视为核心点。核心点及其可达点构成一个簇：

$$\sum_{q \in N_\varepsilon(p)} E(q) \geqslant MinEnergy, \tag{2}$$

式中：$E(q)$表示点 q 的能量值。

3) 可达点：如果一个点在半径 ε 范围内点的能量总和少于 *MinEnergy*，但至少包含一个核心点，该点则称为可达点。它可以通过连接核心点被包含在一个簇中。

4) 离群点：如果一个点在半径 ε 范围内点的能量总和少于 *MinEnergy* 且不包含任何核心点，则该点被标记为离群点。离群点不包含在任何簇中。

对于一个点集，将能量总和超过某个阈值的点聚类成一组，而低于该阈值的也未被任何核心点邻域覆盖的点被视为噪声点，使聚类结果更偏向高能量区域。对于频率—速度域的面波频散谱，每个数据点都包含代表波传播物理能量的数值，该算法考虑到了离散点的能量，最终能够从频散谱中区分不同模式的区域。

为说明 E-DBSCAN 算法的聚类特性，构建了一组合成散点数据集。该数据集包含 6 条能量带，共 3 500 个能量点，其中有两组能量带相距较近。如图 1 所示，分别采用传统 DBSCAN 以及 E-DBSCAN 算法对该数据集进行聚类处理，其中 X、Y 表示散点欧氏距离。

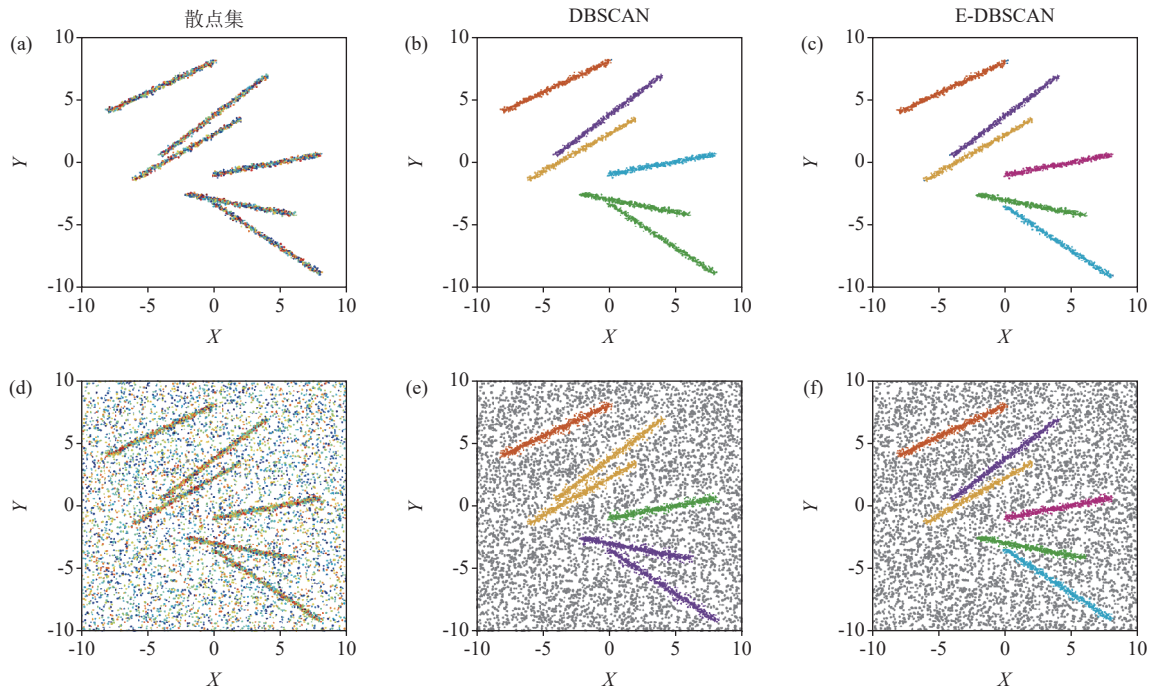
在无噪声条件下，传统 DBSCAN 算法由于基于离散点的密度处理二维数据和聚类，仅将能量点识别并聚类为 5 组。而 E-DBSCAN 算法考虑到了离散点的能量，能够准确区分并识别 6 条能量带，表现出更好的聚类准确性。

进一步在散点集中引入 7 000 个背景噪声点，以模拟实际频散谱中普遍存在的噪声干扰。在含噪情况下，DBSCAN 算法聚类结果出现相互合并的现象，将能量点识别为 4 组。而 E-DBSCAN 算法在含噪情况下仍能以较高准确性区分不同能量带。

1.2 拾取频散曲线的算法设计

E-DBSCAN 算法拾取频散曲线的流程如图 2 所示，具体步骤如下：

第一步：通过频散成像方法将地震数据转换为频散图像。在频散成像过程中，频率和相速度的采样



a—能量点集; b—DBSCAN 算法的聚类结果; c—E-DBSCAN 算法的聚类结果; d—加噪后的能量点集; e—DBSCAN 对加噪数据的聚类结果; f—E-DBSCAN 对加噪数据的聚类结果

a—energy point set; b—clustering result of DBSCAN algorithm; c—clustering result of E-DBSCAN algorithm; d—energy point set with noise; e—clustering result of DBSCAN algorithm for noised data; f—clustering result of E-DBSCAN algorithm for noised data

图 1　传统 DBSCAN 算法与 E-DBSCAN 算法的处理结果对比

Fig. 1　Comparison of clustering results between conventional DBSCAN and E-DBSCAN

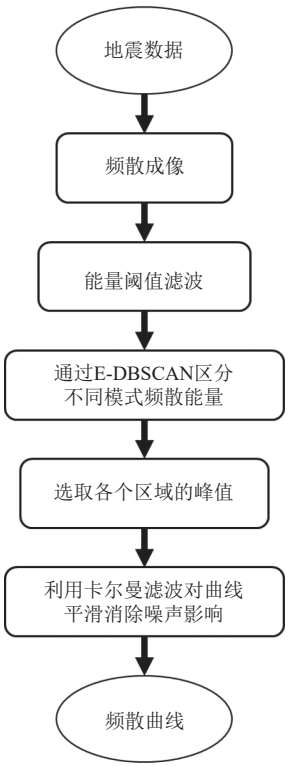


图 2　拾取频散曲线流程

Fig. 2　The flowchart for picking dispersion curves

是离散的。每一对频率和相速度都对应一个能量值。

第二步: 输入组合后的频散谱, 包含坐标(本文中为频率—速度域坐标)及对应的能量。

对频散谱进行能量阈值滤波处理, 去除一部分背景噪声的干扰, 得到滤波后的图像。

第三步: 使用 E-DNSCAN 算法遍历频散谱中的每个点, 得到在频率—速度域中的能量聚类。

第四步: 分别提取每个能量聚类中不同频率的峰值, 并进行卡尔曼滤波平滑这些峰值曲线, 以消除噪声的干扰, 最后得到拾取的频散曲线。

2　模型测试

为了验证所提方法拾取频散曲线的有效性, 本文通过对两个模型的合成地震记录进行测试分析。采用两个层状模型, 通过交错网格有限差分法模拟地震记录, 时间上采用二阶精度, 空间上采用十阶精度, 在左、右和底边界设置完全匹配层吸收边界条件, 吸收层厚度为 $60\Delta x$, 使用主频 25 Hz 的雷克子波垂直力源激发地震波。

2.1　模型 1

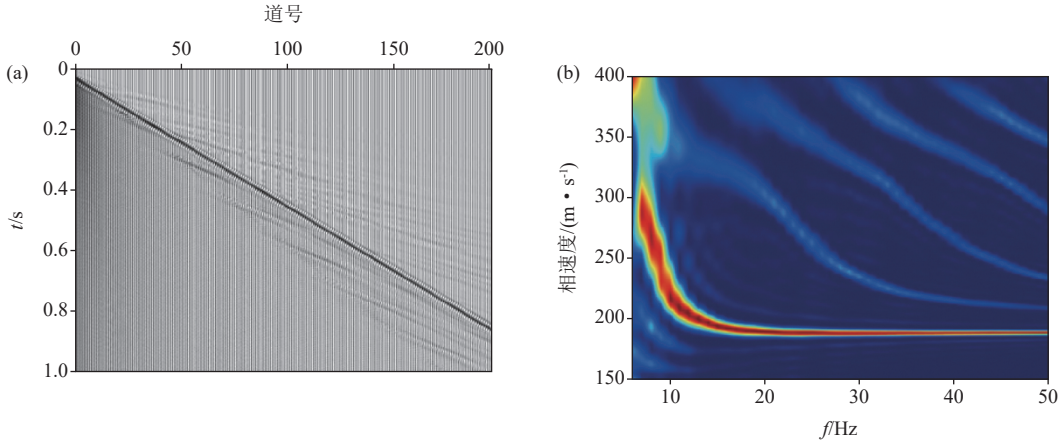
模型 1 的具体参数如表 1 所示。采样间隔为 0.05 ms, 采样时间为 1 s, 道间距为 1 m, 排列长度为 200 道。模拟地震记录如图 3a 所示。通过相移法生成频散谱, 如图 3b 所示。

由于频散能量图中常混有噪声或其他无关信号，基于能量峰值直接拾取频散曲线的方法缺乏针对性，难以有效排除干扰，会导致拾取的频散曲线存在连续性缺失及平滑性不足等问题，拾取的频散曲线不准确。针对上述问题，如图 4 所示，以模型 1 的结果为例说明本文方法的频散曲线拾取流程。首先利用能量阈值滤波对图 3b 所示的频散谱进行处理，过滤掉低振幅的背景噪声，结果如图 4a 所示。然后采用 E-DBSCAN 算法对能量阈值滤波后的频散谱

表 1　模型 1 弹性参数

Table 1　Elastic parameters of model 1

层序	层厚 /m	v_P/ (m·s^{-1})	v_S/ (m·s^{-1})	ρ/ (kg·m^{-3})	泊松比
1	10	400	200	2000	0.33
2	∞	800	400	2000	0.33

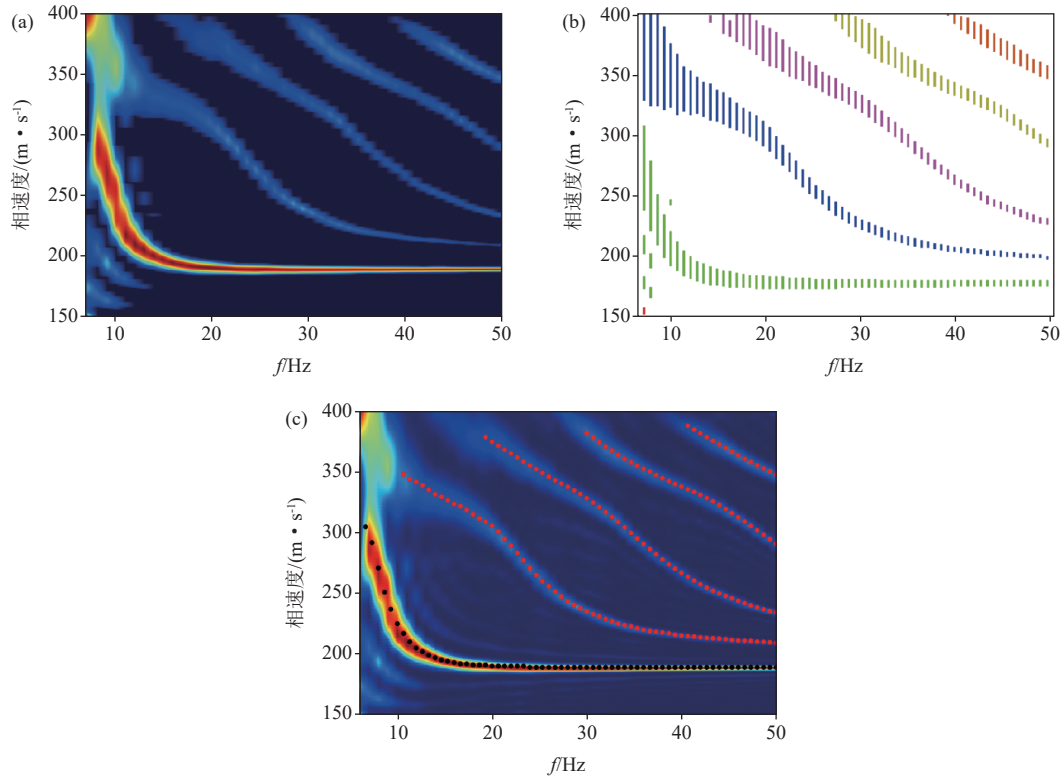


a—模型 1 地震记录; b—模型 1 频散谱

a—seismic record of model 1; b—dispersion spectrum of model 1

图 3　模型 1 数值模拟结果

Fig. 3　Numerical simulation results of mode l



a—能量阈值滤波结果; b—能量密度聚类结果; c—拾取的频散曲线(虚线)

a—energy threshold filtering result; b—energy density clustering result; c—picked dispersion curves (dashed lines)

图 4　模型 1 频散曲线自动拾取过程

Fig. 4　Automatic picking process of dispersion curves for model 1

进行聚类 (图 4b), 区分不同的面波模式。最后, 在不同面波模式的区域中, 对于每个频率, 我们提取不同相速度对应的最大振幅值。然后, 将同一模式区域中的最大振幅值连接起来, 得到频散曲线。在处理野外数据时, 频散图像中的噪声也会影响频散曲线的提取, 尤其是对于信噪比较低的野外数据。为了进一步消除噪声的干扰, 可以使用卡尔曼滤波对初步提取的频散曲线进行平滑处理。经过处理后的频散曲线会更加平滑和合理。拾取效果如图 4c 虚线所示。

将自动拾取的频散曲线与理论频散曲线对比, 结果如图 5 所示(其中黑色实线为理论频散曲线, 红色虚线为自动拾取的频散曲线)。可以看出, 该方法拾取的频散曲线与理论频散曲线能够较好地匹配。采用平均相对误差(mean relative error, MRE)来定量分析自动拾取的频散曲线的准确性。MRE 由式 (3) 计算得到:

$$MRE = \frac{\sum_{i=1}^{N}\left|\frac{R_i - T_i}{T_i}\right|}{N} \times 100\% , \qquad (3)$$

式中: N 表示点数, R_i 和 T_i 分别表示拾取频散曲线和理论频散曲线。由式 (3) 得, 二者平均误差为 0.13%。

2.2　模型 2

模型 2 具体参数如表 2 所示。采样间隔为 0.1 ms, 采样时间为 0.5 s, 道间距为 1 m, 排列长度为 200 道。模拟地震记录如图 6a 所示。通过相移法生成频散谱, 如图 6b 所示。图 7a 是能量阈值滤波结果, 图 7b 是能量密度聚类结果。在不同面波模式的区域中, 对于每个频率提取峰值, 平滑后实现频散曲线的自动拾取, 拾取效果如图 7c 黑色和红色虚线所示。

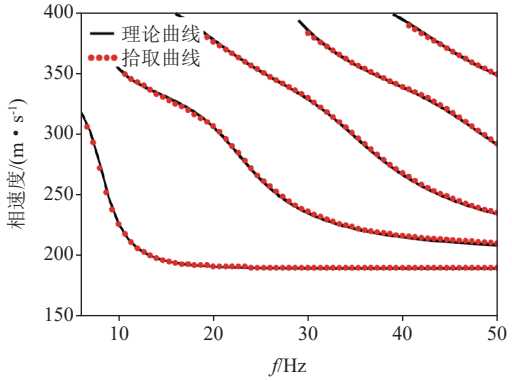


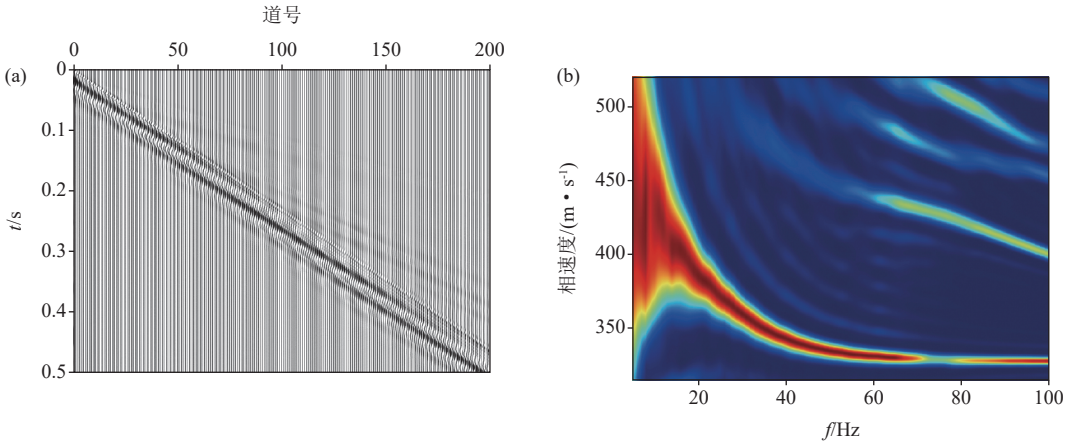
图 5　模型 1 频散曲线拾取结果对比

Fig. 5　Comparison of dispersion curves picking results for model 1

表 2　模型 2 弹性参数

Table 2　Elastic parameters of model 2

层序	层厚 /m	v_P/ (m·s^{-1})	v_S/ (m·s^{-1})	ρ/ (kg·m^{-3})	泊松比
1	5	633	350	2000	0.28
2	10	766	450	2000	0.24
3	∞	859	520	2000	0.21



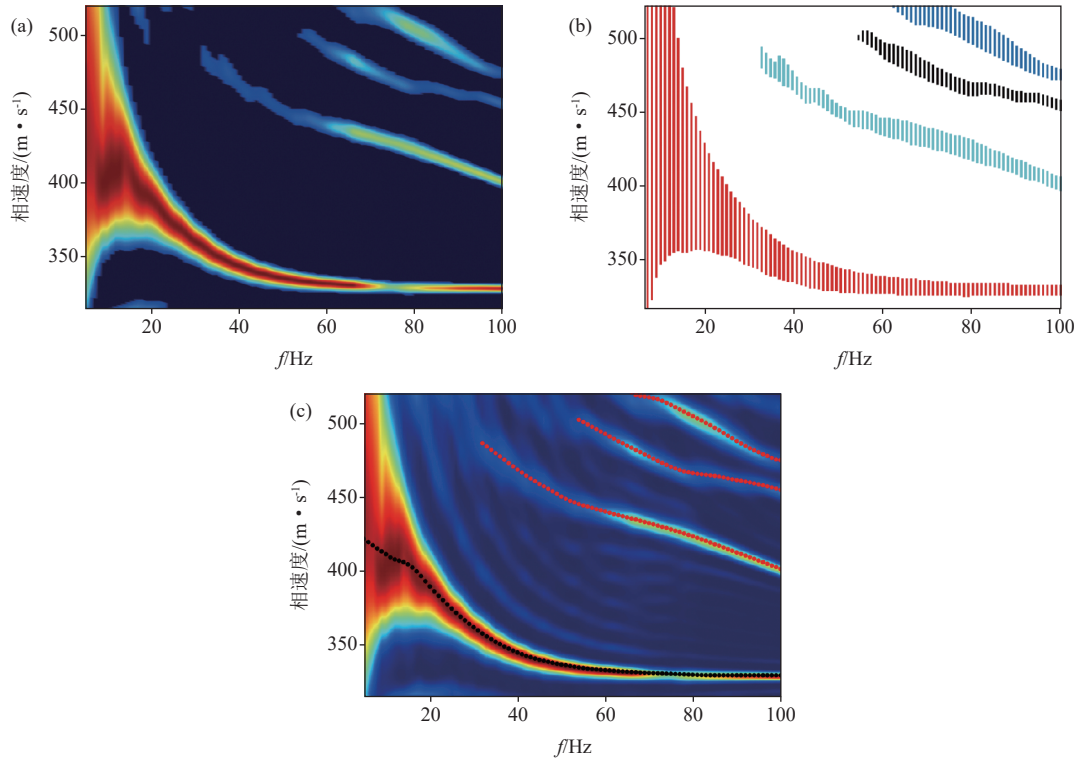
a—模型 2 地震记录; b—模型 2 频散谱

a—seismic record of model 2; b—dispersion spectrum of model 2

图 6　模型 2 数值模拟结果

Fig. 6　Numerical simulation results of model 2

将拾取频散曲线与理论频散曲线对比, 结果如图 8 所示, 图 8 中黑色实线为理论频散曲线, 红色虚线为拾取的频散曲线, 二者的平均相对误差为 2.06%。拾取的多阶频散曲线与理论值基本吻合。基阶和第一高阶的低频段误差相对较大, 这主要源于有限差分法的数值误差以及阵列长度不足以使面波低频充分发育。结果显示, 本文方法拾取的频散曲线与理论频散曲线二者匹配程度较高, 进一步



a—能量阈值滤波结果; b—能量密度聚类结果; c—拾取的频散曲线(虚线)

a—energy threshold filtering result; b—energy density clustering result; c—picked dispersion curves(dashed lines)

图 7　模型 2 频散曲线自动拾取过程

Fig. 7　Automatic picking process of dispersion curves for model 2

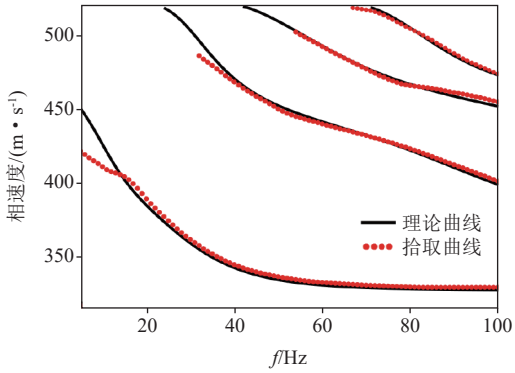


图 8　模型 2 频散曲线拾取结果对比

Fig. 8　Comparison of dispersion curves picking results for model 2

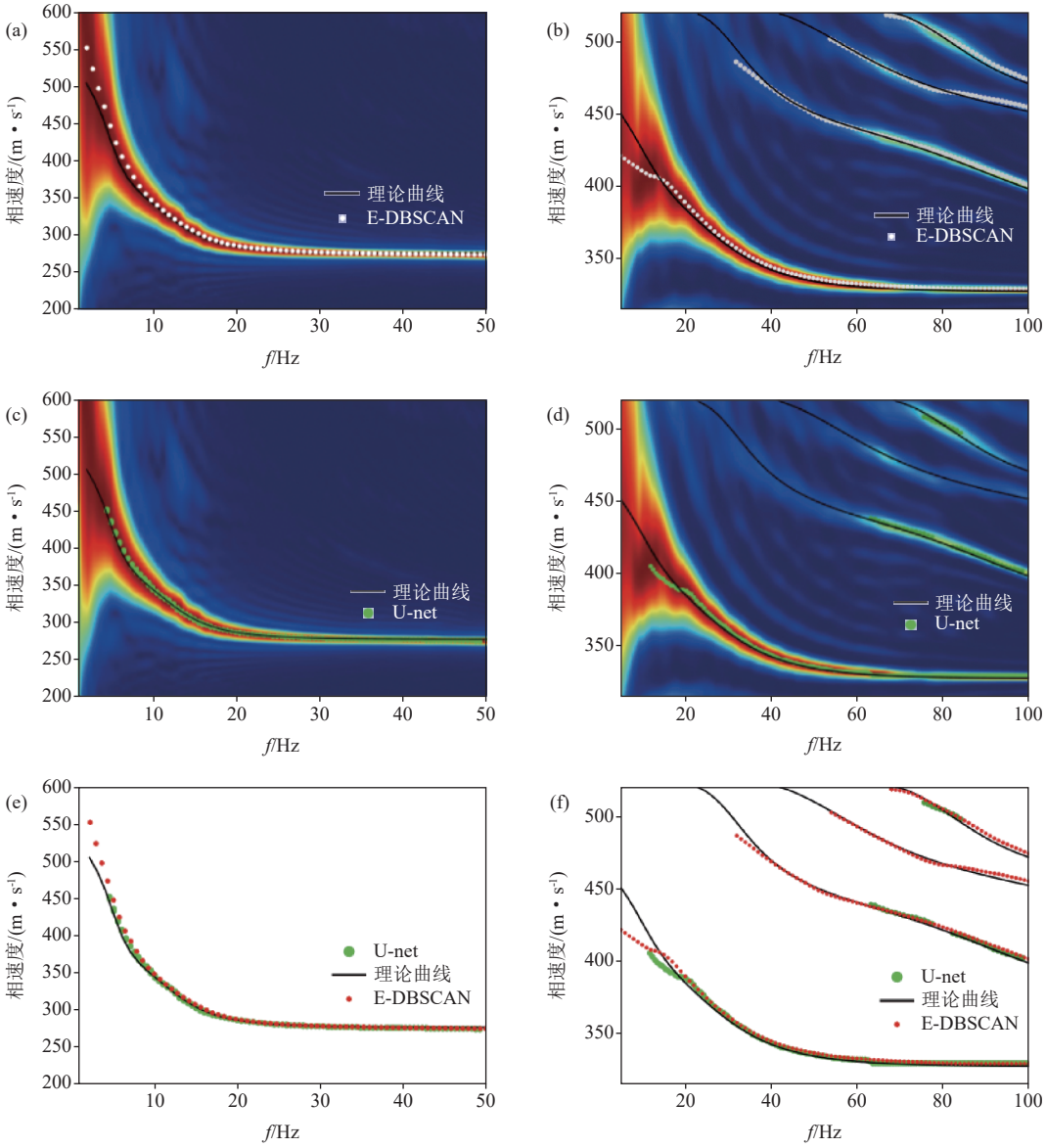
证明应用 E-DBSCAN 算法自动拾取频散曲线的可靠性。

2.3　方法对比

为比较不同方法在不同模型下的频散曲线拾取能力, 我们将所提出方法与基于深度学习的频散曲线自动拾取方法进行了对比分析。对比方法采用 U-net 网络结构, 包含 5 层下采样与 5 层上采样, 以频散谱作为输入, 对应的频散曲线作为输出。训练数据由合成地震记录的频散谱及理论频散曲线作为标签构成, 共建立 100 组样本数据集。网络迭代训练 200 次, 在 NVIDIA 5060Ti GPU 上总耗时 52 min。

图 9a, b 为本文方法的拾取结果, 图 9c, d 为 U-net 方法的拾取结果, 图 9e, f 为拾取曲线与理论曲线的对比结果。对于第一个频散谱, 本文方法的拾取耗时为 1.06 s, U-net 方法的拾取耗时为 2.62 s; 对于第二个频散谱, 本文方法的拾取耗时为 1.21 s, U-net 方法的拾取耗时为 3.70 s。

在拾取效率方面, 本文方法无需预训练数据, 仅依赖数据内部的分布特征进行处理, 不需要构建训练数据集及进行长时间网络训练, 在保证拾取精度的同时显著降低了计算成本。在拾取准确性方面, 从图 9 第一列可以看出, 在仅包含基阶频散能量的情况下, 两种方法均能够较好地贴合理论频散曲线; 而在包含多阶频散能量的条件下, 本文方法仍能够较为准确地拾取高阶频散曲线, 体现出良好的适用性, 而 U-net 方法在高阶频散能量区域出现一定程度的误拾和漏拾现象。在低频部分(0~10 Hz), 拾取频散曲线与理论频散曲线存在较大偏差。这主要源于正演模拟的数值误差以及排列长度不足, 使面波低频发育不充分, 进而导致低频频散能量聚焦性降低, 能量峰值相对于理论频散曲线发生偏移。综合对比结果表明, 本文所提出的方法能够在更短时间内获得更可靠的多阶频散曲线。



a—模型 1E-DBSCAN 拾取结果; b—模型 2E-DBSCAN 拾取结果; c—模型 1U-net 拾取结果; d—模型 2U-net 拾取结果; e—模型 1 曲线对比; f—模型 2 曲线对比

a—E-DBSCAN pick result of model 1; b—E-DBSCAN pick result of model 2; c—U-net pick result of model 1; d—U-net pick result of model 2; e—comparison of picking results for spectrum of model 1; f— comparison of picking results for spectrum of model 2

图 9　不同方法拾取频散曲线对比

Fig. 9　Comparisons of the dispersion curves picked by different methods

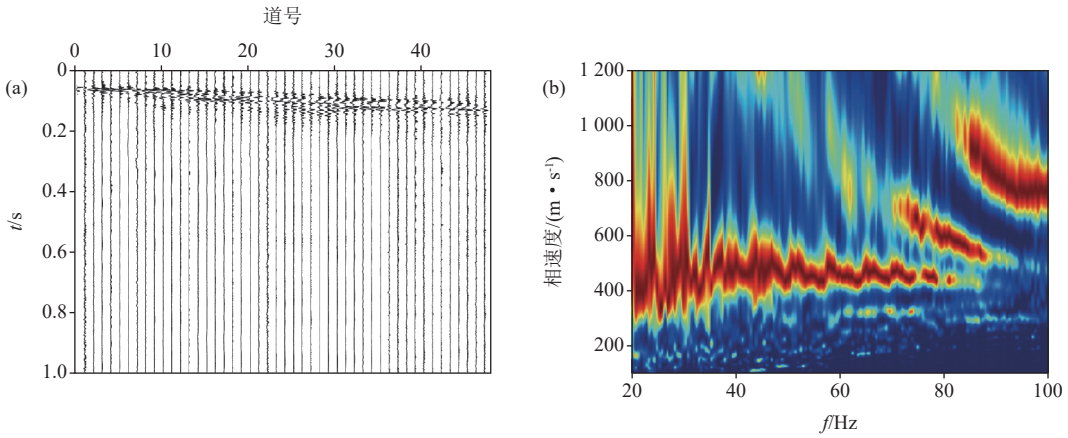
3　实际数据验证

为了验证本文方法处理实际数据的效果, 采用某地采集的地震数据其中一炮进行验证。该数据的最小偏移距为 2.5 m, 道间距 0.5 m, 采样间隔 2 ms, 采样时间 1 s, 共 48 道。其地震记录如图 10a 所示, 通过相移法得到频散谱如图 10b。拾取效果如图 11 所示, 其中黑色虚线表示利用本文方法自动拾取的频散曲线, 可以看出本文方法自动拾取的频散曲线与频散谱中的能量峰值较为匹配且光滑平整, 表明了该方法的可靠性。

为进一步评估本文方法在实际工区数据处理中的计算效率, 采用该自动拾取方法对 70 组频散成像结果进行了批量处理。在戴尔 OptiPlex 7000 计算机(英特尔酷睿 i7-12700 2.1 GHz 处理器 16G 内存)上总耗时为 49 s, 相较于传统人工拾取的方式能够显著缩短处理时间, 满足大规模数据快速处理的需求。

4　结论

本文基于 E-DBSCAN 算法形成了一种多阶面



a—地震记录; b—频散谱

a—seismic record; b—dispersion spectrum

图 10　实际地震记录及其频散谱

Fig. 10　Field seismic data and dispersion spectrum

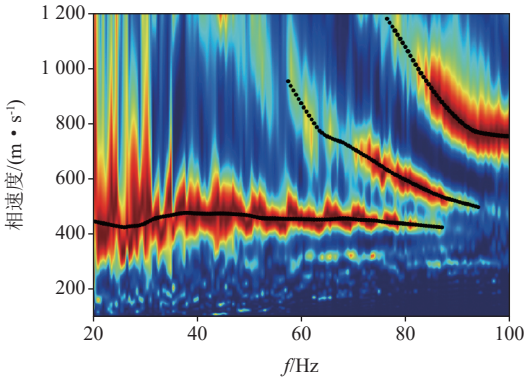


图 11　实际数据频散曲线拾取结果

Fig. 11　Dispersion curves picking results for field data

波频散曲线自动拾取方法。该方法能够从复杂频散谱中准确区分面波模式并拾取连续、稳定的多阶频散曲线，在一定程度上提高了频散分析的处理效率。通过两组合成模型以及实际数据进行了测试，并与其他机器学习方法进行对比，结果表明本文方法在拾取精度、连续性及抗噪性能方面优于传统方法，验证了其有效性与可行性。

需要指出的是，频散曲线自动拾取的精度主要受频散成像质量的影响。当信噪比较低时，聚类边界可能不够清晰，从而影响拾取结果。此外，低频频散能量受观测尺度及采集参数限制较为明显，可能导致频散曲线在低频段出现一定偏差。因此在关键频段或能量分布异常区域，仍可结合人工经验进行调整，以确保拾取结果的可靠性。

由于算法参数的选择取决于频散谱的质量。因此对于不同的数据集，需要在少量数据上进行测试。未来可进一步研究聚类参数的自适应选取策略，使算法能够根据频散能量特征自动调整关键参数，从而减少人工干预，提高方法在不同条件下的普适性。

参考文献(References)：

[1] Park C, Miller R, Xia J. Imaging dispersion curves of surface waves on multi-channel record[C]//SEG Technical Program Expanded Abstracts 1998,Society of Exploration Geophysicists, 1998: 1377-1380.

[2] Luo Y H, Xia J H, Miller R D, et al. Rayleigh-wave dispersive energy imaging using a high-resolution linear radon transform[J]. Pure and Applied Geophysics, 2008, 165(5): 903−922.

[3] 伍敦仕, 孙成禹, 林美言. 基于频率-速度域多重信号分类的面波高分辨率频散成像方法 [J]. 石油物探, 2017, 56(1): 141−149.
Wu D S, Sun C Y, Lin M Y. High resolution dispersion imaging of surface waves based on multiple signal classification in frequency-velocity domain[J]. Geophysical Prospecting for Petroleum, 2017, 56(1): 141−149.

[4] McMechan G A, Yedlin M J. Analysis of dispersive by wave field transformation[J]. Geophysics, 1981, 46(6): 869−874.

[5] Xia J H, Xu Y X, Miller R D. Generating an image of dispersive energy by frequency decomposition and slant stacking[J]. Pure and Applied Geophysics, 2007, 164(5): 941−956.

[6] 伍敦仕, 孙成禹, 林美言. 基于互相关相移的主动源地震面波频散成像方法 [J]. 地球物理学进展, 2017, 32(4): 1693−1700.
Wu D S, Sun C Y, Lin M Y. Activeseismic surface wave dispersion imaging method based on cross-correlation and phase-shifting[J]. Progress in Geophysics, 2017, 32(4): 1693−1700.

[7] Granados I, Calò M, Ramos V. Noisy dispersion curve picking (NDCP): A Matlab package for group velocity dispersion picking of seismic surface waves[J]. Computers & Geosciences, 2019, 133: 104315.

[8] Bergen K J, Chen T, Li Z F. Preface to the focus section on machine learning in seismology[J]. Seismological Research Letters, 2019, 90(2A): 477−480.

[9] Al-Mashhor A A, Al-Shuhail A A, Hanafy S M, et al. First arrival picking of seismic data based on trace envelope[J]. IEEE Access, 2019, 7: 128806−128815.

[10] Hong Z, Su M J, Qian F, et al. Global seismic horizon interpretation based on data mining:A new tool for seismic geomorphologic study[J]. Interpretation, 2020, 8(1): T131−T140.

[11] Cárdenas J, Burzawa A, Radic N, et al. Deep learning contribution to the automatic picking of surface-wave dispersion for the charac-

terization of railway earthworks[J]. Computers & Geosciences, 2025, 198: 105883.

[12] Ortega R, Carciumaru D, Huesca E, et al. Automatic selection of dispersion curves based on a weighted probability scheme[J]. Seismological Research Letters, 2019, 90(2A): 623−632.

[13] Xie P, Boelle J L, Khosro F, et al. Automatic surface wave dispersion curve picking and symbolic calculation inversion[C]//79th EAGE Conference and Exhibition 2017, 2017.

[14] Taipodia J, Dey A, Gaj S, et al. Quantification of the resolution of dispersion image in active MASW survey and automated extraction of dispersion curve[J]. Computers & Geosciences, 2020, 135: 104360.

[15] Rovetta D, Kontakis A, Colombo D, et al. A density-based spatial clustering application for a fully automatic picking of surface wave dispersion curves[C]//Virtual:SEG Technical Program Expanded Abstracts 2020,Society of Exploration Geophysicists, 2020: 1850-1854.

[16] Wang Z N, Sun C Y, Wu D S. Automatic picking of multi-mode surface-wave dispersion curves based on machine learning clustering methods[J]. Computers & Geosciences, 2021, 153: 104809.

[17] Liu H, Li J, Hu R. Automatic and adaptive picking of surface-wave dispersion curves for near-surface application[J]. Journal of Applied Geophysics, 2024, 221: 105282.

[18] Yao H, Cao W P, Huang X R, et al. Automatic extraction of surface wave dispersion curves using unsupervised learning and high-resolution tau-p transform[J]. Earth and Space Science, 2023, 10(12): e2023EA003198.

[19] 康娜, 张薇薇, 曹静杰. 基于区域生长算法的频散曲线自动拾取 [J]. 工程地球物理学报, 2025, 22(4): 586−594.
Kang N, Zhang W W, Cao J J. Automatic picking of dispersion curves based on region growing algorithm[J]. Chinese Journal of Engineering Geophysics, 2025, 22(4): 586−594.

[20] Kuang X J, Pan Y D, Zhang Z Y, et al. Automatic picking of multimodal Rayleigh-wave dispersion curves from multicomponent data with an energy-density-based clustering method[J]. Geophysical Journal International, 2025, 243(1): ggaf323.

[21] 陈杰, 熊章强, 张大洲, 等. 基于多模式的多重滤波方法提取瑞雷面波频散曲线 [J]. 工程地球物理学报, 2018, 15(4): 411−417.
Chen J, Xiong Z Q, Zhang D Z, et al. The extract dispersion curves of Rayleigh surface wave based on multiple filtering method[J]. Chinese Journal of Engineering Geophysics, 2018, 15(4): 411−417.

[22] 宋欣悦, 刘国昌, 杜婧, 等. 基于面波频散曲线聚类分析的近地表横波速度反演 [J]. 地球物理学报, 2023, 66(7): 3026−3047.
Song X Y, Liu G C, Du J, et al. Near surface S-wave velocity inversion based on cluster analysis of surface wave dispersion curve[J]. Chinese Journal of Geophysics, 2023, 66(7): 3026−3047.

[23] Alyousuf T, Colombo D, Rovetta D, et al. Near-surface velocity analysis for single-sensor data: An integrated workflow using surface waves, AI, and structure-regularized inversion[C]// Anaheim Califo: SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists, 2018: 2342-2346.

[24] Dai T Y, Xia J H, Ning L, et al. Deep learning for extracting dispersion curves[J]. Surveys in Geophysics, 2021, 42(1): 69−95.

[25] 伍锡军. 基于深度学习的高频面波反演方法研究 [D]. 成都: 电子科技大学, 2020.
Wu X J. Research on inversion method of high frequency surface wave based on deep learning[D]. Chengdu: University of Electronic Science and Technology of China, 2020.

[26] Zhang X T, Jia Z, Ross Z E, et al. Extracting dispersion curves from ambient noise correlations using deep learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(12): 8932−8939.

A method for automatically picking surface wave dispersion curves based on the energy density-based spatial clustering of applications with noise algorithm

ZHAO Yun-Long, SUN Cheng-Yu, HUANG Hong-Yu

(*School of Geosciences*, *China University of Pertoleum* (*East China*), *Qingdao*　266580, *China*)

Abstract: Surface wave analysis, used to characterize near-surface shear-wave velocity structures, has been widely applied to various distinct fields. Extracting surface wave dispersion information serves as a significant step in surface wave analysis. Conventional methods for picking dispersion curves primarily include manual picking and semi-automatic picking in manually delineated dispersion energy zones. Given the large volume of seismic data used in oil and gas exploration, manual processing methods struggle to satisfy the requirements of practical production. Therefore, this study proposes a method for automatically picking multi-mode surface-wave dispersion curves based on an unsupervised machine learning approach. First, some background noise points in a dispersion spectrum are removed using energy threshold filtering. Second, different surface wave modes in the dispersion spectrum are determined using the energy density-based spatial clustering of applications with noise (E-DBSCAN) algorithm. Third, energy peaks in zones corresponding to different surface wave modes are searched, followed by noise elimination through Kalman filtering and smoothing. As a result, multi-mode dispersion curves are obtained. The proposed method incorporates dispersion energy values, offering high reliability. Numerical experiments demonstrate that the automatically picked dispersion curves accurately match the theoretical dispersion curves. Meanwhile, method comparison and field data tests further confirm the reliability of the proposed method.

Key words: surface wave; dispersion curve; clustering algorithm; automatic picking

(本文编辑: 叶佩)