



## 数学形态学和 K-SVD 字典学习 在大地电磁数据去噪中的应用

桂团福<sup>1,2,3</sup>, 邓居智<sup>1,3</sup>, 李 广<sup>1,2,3</sup>, 刘晓琼<sup>1,2</sup>, 陈 辉<sup>1,3</sup>, 何柱石<sup>1,2,3</sup>

- (1. 江西省防震减灾与工程地质灾害探测工程研究中心(东华理工大学), 南昌 330013;
2. 江西省放射性地学大数据技术工程实验室(东华理工大学), 南昌 330013;
3. 东华理工大学 地球物理与测控技术学院, 南昌 330013)

**摘 要:**为解决大地电磁(Magnetotelluric, MT)常规的时间域去噪方法对于 1 Hz 附近噪声压制的局限性, 提出了一种基于数学形态滤波(Mathematical morphological filtering, MMF)和 K-SVD(K-Singular value decomposition)字典学习的新型去噪方法, 用于压制 MT 信号中低频数据 1Hz 附近的强人文噪声。首先, 利用 MMF 分离出低频信号, 对该低频信号进行保护以防止信号丢失; 然后, 使用 K-SVD 字典学习对剩余的含噪信号进行处理, 通过从观测数据中自主学习获取噪声的特征结构, 提取噪声轮廓, 达到去除噪声的目的。利用一个合成数据集验证算法后, 对两个实测数据进行处理, 结果表明: 该方法可以在几乎不损失有效信号的前提下, 消除各种强人文噪声, 信噪比大幅提升, 数据质量得到很大改善, 且去噪效果优于小波变换等传统方法。

**关键词:** 大地电磁; 数学形态滤波; K-SVD 字典学习; 强干扰压制; 低频信号

**文章编号:** 1004-0609(2021)-12-3713-17

**中图分类号:** P631

**文献标志码:** A

**引文格式:** 桂团福, 邓居智, 李 广, 等. 数学形态学和 K-SVD 字典学习在大地电磁数据去噪中的应用[J]. 中国有色金属学报, 2021, 31(12): 3713–3729. DOI: 10.11817/j.ysxb.1004.0609.2021-40078

GUI Tuan-fu, DENG Ju-zhi, LI Guang, et al. De-noising magnetotelluric data based on mathematical morphology and K-SVD dictionary learning[J]. The Chinese Journal of Nonferrous Metals, 2021, 31(12): 3713–3729. DOI: 10.11817/j.ysxb.1004.0609.2021-40078

大地电磁测深法(MT)是常用的地球深部探测方法之一, 它具有探测深度大、施工简便、不受高阻层屏蔽、受地形制约较小等诸多优点, 已在矿产资源勘查、能源探测、地球深部电性结构研究等领域得到广泛应用<sup>[1-2]</sup>。但是, MT 利用的天然电磁信号随机性强、频带宽且幅值较小, 极易受到人文噪声的干扰。噪声的存在对于反演的结果有很大影响, 显著降低探测结果的分辨率<sup>[3]</sup>, 强人文噪声已成为 MT 方法应用的主要瓶颈之一<sup>[4]</sup>。

为了提高 MT 数据的信噪比, 前人已开发了多

种去噪方法, 其中远参考方法<sup>[5]</sup>和 Robust 估计方法<sup>[6]</sup>是最常用且广为认可的方法, 被认为是大地电磁信号处理的两个里程碑<sup>[7]</sup>。但是, 远参考方法的结果高度依赖于在远参考站采集的数据质量。Robust 估计的效果依赖于观测数据中高质量的数据占主导地位; 当受到持续性的强人文噪声污染时, Robust 估计方法很难取得较好的效果<sup>[8]</sup>。基于视电阻率-相位的转换关系以及数据的连续性, 利用高质量的视电阻率或者相位数据, 对测深曲线进行校正也是一种可行的方法, 包括 Rhoplus 及 Dplus 校正法等<sup>[9]</sup>。

**基金项目:** 国家自然科学基金资助项目(41904076, 41830107, 41674077); 国家重点研发计划资助项目(2018YFC0603202); 江西省自然科学基金资助项目(20192BAB212009); 江西省放射性地学大数据技术工程实验室开放基金资助项目(JELRBDT202003); 江西省防震减灾与工程地质灾害探测工程研究中心开放基金资助项目(SDGD202008)

**收稿日期:** 2020-12-17; **修订日期:** 2021-04-25

**通信作者:** 邓居智, 教授, 博士; 电话: 13507948149; E-mail: jzdeng73@gmail.com

但在实际情况中,强人文噪声往往造成大量频点畸变,此时采用 Rhoplus 等校正法可能适得其反。基于站间天然电磁场单位脉冲响应的 MT 去噪方法不适用于连续人文噪声干扰的情况。另外,该方法还要保证参考点受噪声干扰较小,其去噪效果与合成数据的精度和噪声识别的精度有关<sup>[10]</sup>。时间域信噪分离最典型的代表为小波变换去噪<sup>[11-12]</sup>,但小波变换极易损失低频信号。独立分量分析也常用于大地电磁去噪,但它需要在有效信号与噪声信号为相互统计独立的信号的前提下,才能取得较好的效果<sup>[13]</sup>。在连续小波变换的基础上,有学者利用信号谱估计方法压制大地电磁噪声,然而对于小波参数的选取需要经过反复尝试<sup>[14]</sup>。许滔滔等<sup>[15]</sup>尝试将机器学习中的神经网络算法应用于 MT 数据的工频干扰压制,但对于其他类型的噪声压制有待进一步研究。上述时域处理方法因直接消除了振幅异常大的噪声均能够提高数据质量,但从时间序列中直接提取噪声极易损失有效的低频信号。因此,这些方法对 MT 数据中 1Hz 以下的低频数据的去噪效果不佳<sup>[16]</sup>。

为了减少低频信号的丢失,汤井田等<sup>[17]</sup>提出了基于稀疏分解和贪婪算法的大地电磁去噪方法,设计了与人文噪声相匹配但对有效信号不敏感、仅处理噪声片段的冗余字典<sup>[18]</sup>。李晋等<sup>[19]</sup>利用多重分形谱将 MT 时间序列分为噪声片段和无噪声片段,然后仅对含噪片段进行处理。尽管这些方法可以有效地保留无噪声段的低频信号,但是难以保留噪声段有用的低频信号。利用多参考站进行阵列 MT 数据处理时,对于 0.1~1 Hz 附近的低频噪声有一定的压制作用,但是需要优选参考站数据,且要求多个站点同步观测<sup>[20]</sup>。

为了解决上述问题,本文将 MMF 与 K-SVD 字典学习有机结合,提出了一种新的强干扰压制方法。该方法首先使用 MMF 提取并保护有效的低频 MT 信号,然后使用 K-SVD 字典学习自主学习获取噪声的特征结构,构造出噪声轮廓,从而去除噪声。这种方法与其他现有时间序列编辑方法之间的主要区别在于,不直接从原始数据中提取噪声,而是首先提取有效的低频信号部分,可克服常规方法损失低频有效信号的问题。本文通过比较合成数据和实测的 MT 数据,验证了新方法的可行性,表明新方法优于传统的时间序列编辑方法。

## 1 方法与理论

### 1.1 数学形态滤波

数学形态学(Mathematical morphology, MM)是从对二元多孔介质的几何学的研究中得出的一种非线性信号处理方法。该方法是一种从数学形态学发展而来的非线性滤波技术,最早是以图像的形态特征作为研究对象,现已成功应用于图像处理、图形分析、计算机视觉和地震数据噪声抑制<sup>[21]</sup>等工程实践领域。MMF 在数据处理领域应用广泛,在传统形态滤波的基础上提出了组合广义形态滤波,对于 MT 信号中噪声的压制具有一定的效果<sup>[22]</sup>。

MMF 有两个基本运算,即腐蚀和膨胀。腐蚀(膨胀)运算等价于滑动过滤器窗口内离散函数的最小(最大)过滤(等效于结构元素)。腐蚀运算是收缩的过程,可用于消除对象的边界点并减少尖峰。膨胀运算是扩张的过程,当边界不光滑时,可用于填充凹部<sup>[23]</sup>。其定义如下。

假设  $A=[a_1, a_2, \dots, a_N]^T$  为要处理的原始 MT 信号,假设  $B=[b_1, b_2, \dots, b_M]^T$  为结构 100 元素,进行腐蚀和膨胀操作定义为:

$$A \odot B = \min_{m=1,2,\dots,M} \{a(n+m)-b(m)\}, \quad (1)$$

$$n=1,2,\dots,(N+M-1)$$

$$A \oplus B = \max_{m=1,2,\dots,M} \{a(n-m)+b(m)\}, \quad (2)$$

$$n=1,2,\dots,(N-M+1)$$

式中:“T”代表矩阵转置;“ $\odot$ ”表示腐蚀运算;“ $\oplus$ ”表示膨胀运算。

通过腐蚀和膨胀运算可以产生新的运算。打开运算可以通过对目标信号进行腐蚀然后进行膨胀来完成;而关闭运算可以通过先执行膨胀然后进行腐蚀来生成:

$$(A \circ B) = [(A \odot B) \oplus B] \quad (3)$$

$$(A \bullet B) = [(A \oplus B) \odot B] \quad (4)$$

式中:“ $\circ$ ”和“ $\bullet$ ”分别是打开和关闭运算。

打开和关闭运算构成了数学形态学中最基本的滤波器。通过执行打开运算然后执行关闭运算来实现开-闭滤波器(OC);通过执行关闭运算然后执行打开运算来实现闭-开滤波器(CO):

$$(OC) = A \circ B \bullet B \quad (5)$$

$$(CO) = A \bullet B \circ B \quad (6)$$

开-闭滤波器和闭-开滤波器都可以滤除目标信号中的噪声。尽管如此, 开-闭滤波器通常会产生较小的输出, 而闭-开滤波器会放大输出。因此, 通常使用这些输出的平均值来近似原始信号:

$$C = \Phi_{OC(CO)}(b) = [OC(A) + CO(A)]/2 \quad (7)$$

在大地电磁数据处理领域, 数学形态滤波最初用于噪声提取。然而, 低频 MT 数据中的主要成分是有有效的低频部分, 对于低于 1 Hz 的 MT 数据尤其如此, 这导致等式中的输出信号  $C$  主要是低频有效信号(不是噪声), 因此该方法不适用于低频 MT 数据的噪声衰减。在本文中, 输出信号  $C$  代表有效的低频 MT 信号, 通过 MMF 处理原始信号, 分离出有效的低频信号和含噪的高频信号。

## 1.2 K-SVD 字典学习

字典学习可以追溯至 1996 年, OLSHAUSEN 等<sup>[24]</sup>提出了 Sparsenet 字典学习方法, 利用最大似然估计学习字典, 实现了少量原子的稀疏表示。ENGAN 等<sup>[25]</sup>在 Sparsenet 字典学习方法基础上提出了 MOD (Method of optimal directions) 字典学习方法。2006 年, AHARON 等<sup>[26]</sup>又提出了 K-SVD 字典学习方法, 该方法通过交替更新字典和系数, 是目前最为常用的字典学习方法。汤井田等<sup>[27]</sup>和李广等<sup>[28]</sup>将字典学习引入到大地电磁数据处理中, 通过字典学习方法自主学习观测数据中人文噪声的特征结构, 实现高精度的信噪分离。

K-SVD 自学习字典是 K-means 算法的一种泛化, 其本质是在稀疏约束条件的限制下, 通过奇异值分解来交替迭代更新字典和稀疏表示的系数。针对给定的样本数据, 通过构建一个具有  $N$  个训练样本的集合  $\{y_i\}_{i=1}^N$ , 按照稀疏约束条件得到一个适合于样本特征的超完备字典。再通过构建的字典进行稀疏表示, 充分利用样本数据的先验信息, 自适应根据样本进行特征提取, 并通过高效的奇异值分解方法进行字典更新和稀疏系数更新。

给定训练数据  $Y$ ,  $Y$  的每一列  $y$  表示一个训练样本, 矩阵  $Y = \{y_i\}_{i=1}^N$  ( $N \gg K$ ) 为  $N$  个训练样本的

集合, 给定  $x$  表示每一个训练样本  $y$  所对应的稀疏系数向量的集合, 给定矩阵  $D$  表示训练得到的超完备字典, 因此字典学习的过程可用一个优化问题表示:

$$\begin{aligned} \min_{D, X} \{ & \|Y - DX\|_F^2 \} \\ \text{subject to} \quad & \forall i, \|x_i\|_0 \leq T_0 \end{aligned} \quad (8)$$

式中:  $Y = \{y_i\}_{i=1}^N$  ( $N \gg K$ ) 为目标函数;  $D \in \mathbb{R}^{n \times K}$  ( $N \gg K$ ) 为稀疏表示字典;  $X = \{x_i\}_{i=1}^N$  为稀疏表示系数;  $\|\bullet\|_F$  采用 Frobenius 范数估计误差;  $\|\bullet\|_0$  为 0 范数, 即非零元素个数;  $T$  为稀疏系数中非零元素个数的最大值。

## 1.3 K-SVD 字典学习过程

通常可以通过两种方式得到初始字典: 一是根据数据样本初始化得到; 二是给定一个字典进行初始化, 如 DCT 字典。本文根据给定样本, 通过混合矩阵方式得到一个初始化字典。

稀疏编码阶段: 根据初始得到的字典, 使用 OMP 正交匹配算法, 求解每一个训练样本  $y_i$  的稀疏系数向量  $x_i$ :

$$\begin{aligned} \min_{x_i} \{ & \|y_i - Dx_i\|_2^2 \} \\ \text{subject to} \quad & \forall i, \|x_i\|_0 \leq T_0 \end{aligned} \quad (9)$$

字典更新: 固定已知的稀疏系数向量  $x_i$  和字典  $D$ , 更新字典的第  $k$  列  $d_k$ , 令稀疏系数  $x_i$  中与  $d_k$  相乘的第  $k$  行向量为  $x$ , 则样本集的目标函数可以写为:

$$\begin{aligned} \|Y - DX\|_F^2 &= \left\| Y - \sum_{j=1}^K d_j x_j^T \right\|_F^2 = \\ &= \left\| \left( Y - \sum_{j \neq k} d_j x_j^T \right) - d_k x_k^T \right\|_F^2 \end{aligned} \quad (10)$$

式中:  $x_j^T$  为稀疏矩阵  $X$  中的第  $j$  行。

定义  $E_k$  为除第  $K$  个原子以外所有原子产生的误差, 则式(10)可表示为:

$$\|Y - DX\|_F^2 = \|E_k - d_k x_k^T\|_F^2 \quad (11)$$

误差  $E_k$  可以表示为:

$$E_k = Y - \sum_{j \neq k} d_j x_j^T \quad (12)$$

定义  $\omega_k = \{i | 1 \leq i \leq K, x_i^T(i) \neq 0\}$  表示使用原子

$d_k$  的样本信号  $Y$  的索引。为确保结果收敛, 定义  $\Omega_k$  为  $N \times \omega_k$  的矩阵, 其中  $(\omega_k(i), i)$  是非零值, 其余都是零值, 则式(11)等价于:

$$\|E_k \Omega_k - d_k x_T^k \Omega_k\|_F^2 = \|E_k^R - d_k x_R^k\|_F^2 \quad (13)$$

式中:  $x_R^k = x_T^k \Omega_k$  表示  $x_T^k$  去掉零值项的行矢量;  $E_k^R = E_k \Omega_k$  表示系数编码过程中使用的原子  $d_k$  的误差列。

将  $E_k^R$  采用奇异值分解法(SVD)得到如下分解表达式:

$$E_k^R = U \Delta V^T \quad (14)$$

通过分解得到  $U$  的第一列来更新初始字典中的  $d_k$ 。同时, 以矩阵  $V$  的第一列与  $\Delta(1,1)$  的乘积来更替系数表示系数  $x_R^k$ 。更新  $j = j + 1$ , 直到 OMP 稀疏分解达到指定阈值, 停止迭代, 输出信号。

#### 1.4 处理流程

图 1 所示为 MMF-KSVD 的基本处理流程图, 首先, 利用 MMF 将含有噪声的 MT 信号分解为低频有效信号和含噪高频信号, 同时保留低频信号。随后, 通过 K-SVD(即, 使用预定的冗余字典)将含噪高频信号分解为高频有效信号和高频人文噪声。最终, 将低频有效信号和高频有效信号拼接以获得全频带有效 MT 信号(完全无噪声的 MT 信号)。

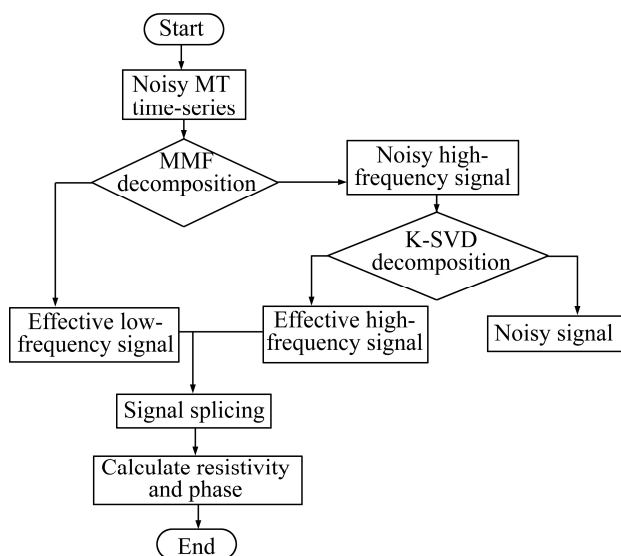


图 1 MMF-KSVD 的基本处理流程

Fig. 1 Basic processing flow of MMF-KSVD

## 2 仿真分析

### 2.1 时间域波形分析

为定量分析所提出方法的有效性, 从青藏高原东北部柴达木盆地收集的大量实测 MT 数据中选择了高质量 MT 数据集。MT 数据采集利用的是 MTU-5A 大地电磁仪。合成数据是通过将模拟噪声添加到良好真实的 TS5 数据集中获得的, 该数据集以 15 Hz 的采样率记录。此外, 使用不同的方法可以降低合成数据集的噪声。选取点 QH301603 噪声较少的信号片段进行仿真测试, 选取该点的其中一段如图 2 所示, 在  $E_x$ 、 $E_y$ 、 $H_x$ 、 $H_y$  方向几乎没有噪声。显然, 在此站点记录的 MT 时间序列是不规则的、随机的, 看起来像是非平稳噪声, 没有异常大的振幅分量; 大多数信号的幅度会随时间缓慢变化。总体而言, 在 QH301603 站点记录的 MT 数据质量高。另外, 在每个通道中, 尤其是在  $E_y$  的电通道中, 都可以找到非常低频的分量。但是, 使用传统的时间序列编辑方法很难将人文噪声与有效的低频成分分离。为了验证方法的可行性, 利用两个电道方向  $E_x$ 、 $E_y$  的低频特性进行仿真处理。

在高质量实测点 QH301603 的信号中, 在  $E_x$ 、 $E_y$  方向中加入三个不同脉宽的方波和一个三角波的噪声信号, 使原始高质量信号存在 4 个不同类型、不同脉宽的噪声信号干扰(由于篇幅有限, 只截取其中一部分时间序列)。其中图 3(a)为原始高质量信号的时间序列, 很容易看出明显的低频特征, 而且几乎没有噪声干扰, 质量较高; 图 3(b)为加入的不同脉宽的三角波和方波的噪声信号, 通过将噪声信号叠加到高质量数据的时间序列中, 合成仿真噪声信号; 图 3(c)为合成的含噪声 MT 信号, 对比  $E_y$  方向加噪前后的时间序列, 加噪后信号的幅值、趋势和各项参数产生很大变化。

为了验证本文方法的优越性, 本文对比了两种经典方法—互补集合经验模态分解(Complementary ensemble empirical mode decomposition, CEEMD)和小波变换(Wavelet)—提取出的高质量 MT 信号中的低频信号。图 4 所示为互补集合经验模态分解(CEEMD)、小波变换(Wavelet)和数学形态滤波(MMF)提取的效果。对比发现, 三种不同方法都能够很好保留原始高质量低频信号。在人为加入四种

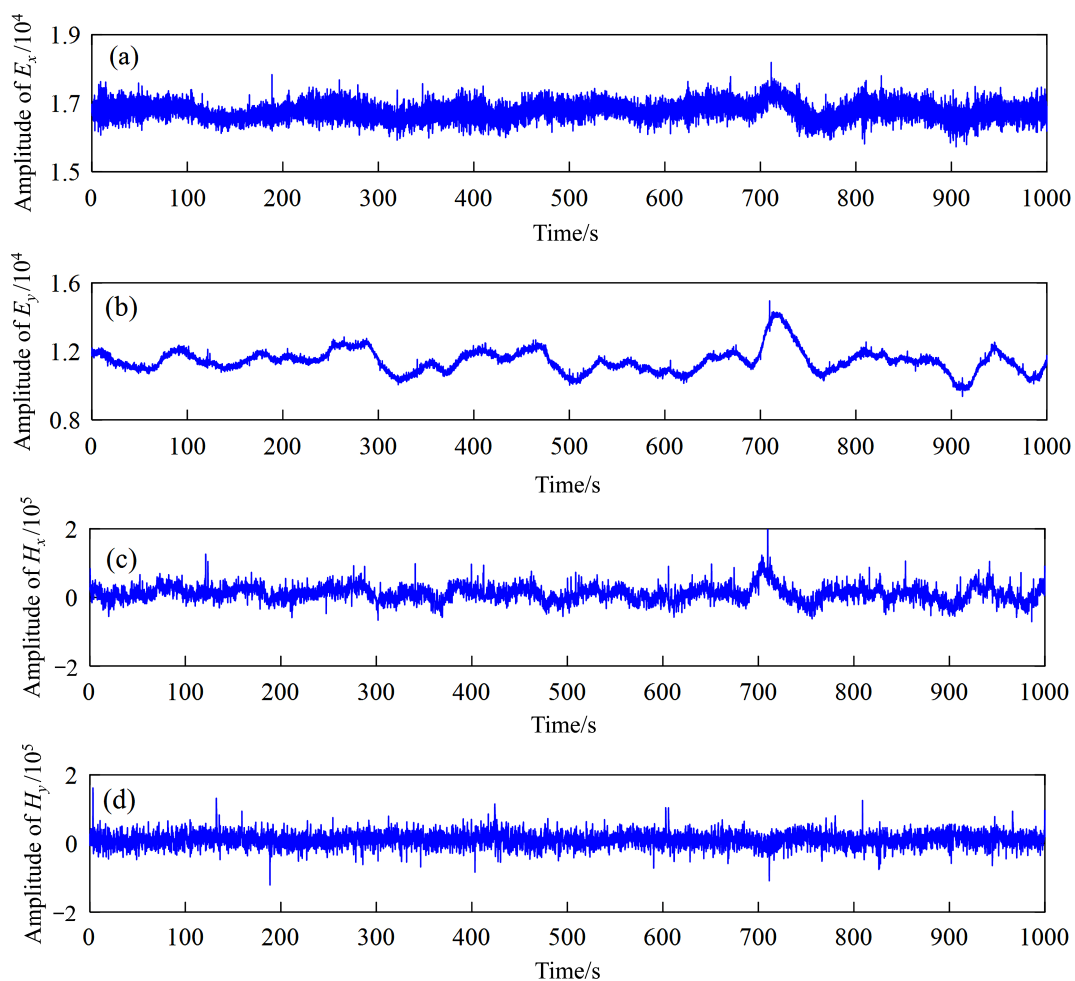


图 2 青海 QH301603 测点  $E_x$ 、 $E_y$ 、 $H_x$ 、 $H_y$  方向的数据

Fig. 2 Time series of  $E_x$  (a),  $E_y$  (b),  $H_x$  (c),  $H_y$  (d) in measured site QH301603 in Qinghai Province, China

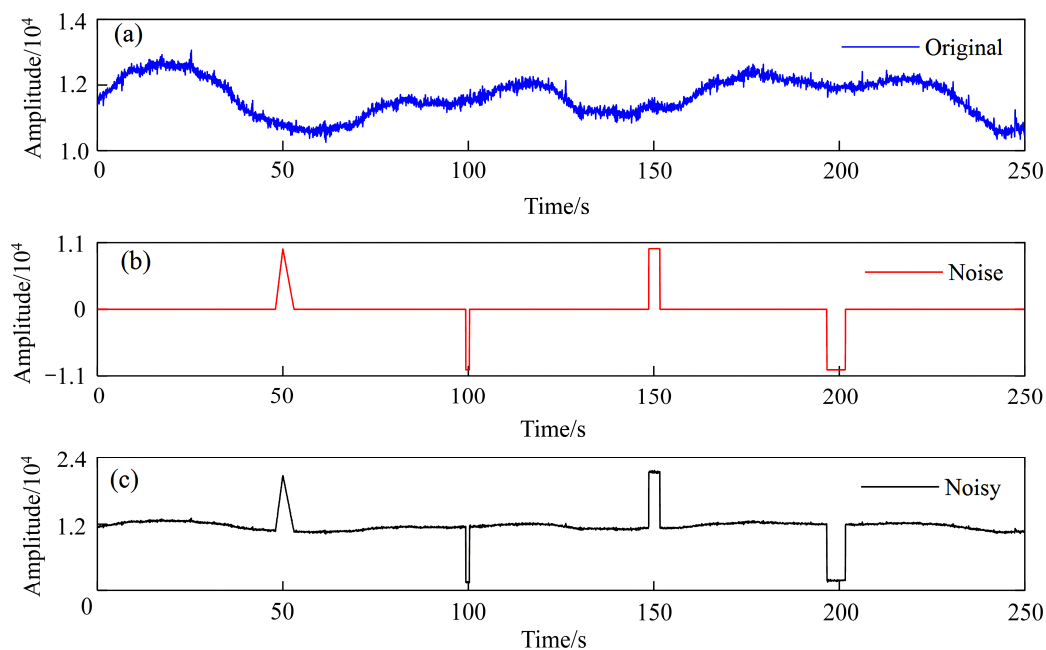


图 3  $E_y$  方向进行加噪处理

Fig. 3 Add noise to time series in  $E_y$  direction: (a) Original; (b) Noise; (c) Noisy

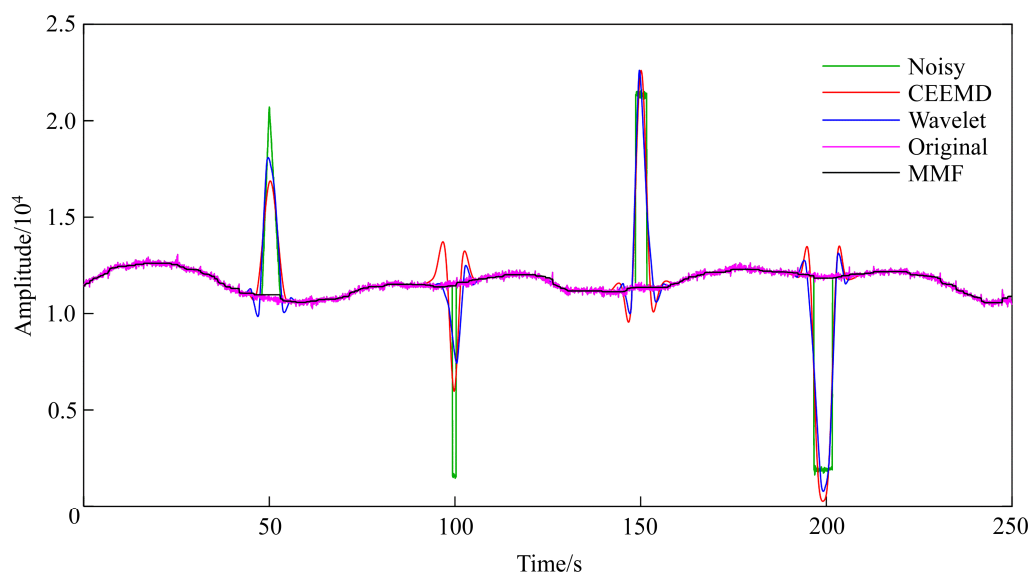


图4 不同方法分离低频和高频分离效果对比

Fig. 4 Comparison of low-frequency and high-frequency separation effects of different methods

仿真噪声后, CEEMD、Wavelet 对于低频信号的提取具有一定的局限性, 结果与原始高质量的低频信号有一定的差距, 而 MMF 可以很好地保留了低频信号。MMF 提取出来的低频信号与原始高质量低频信号基本保持一致。对于仿真信号低频部分的提取, MMF 处理效果具有一定的优越性, 能够保留低频信号。因此, 利用 MMF 能够起到分离低频信号和含噪的高频信号的效果。

结构元素是影响 MMF 处理结果最重要的参数之一。利用圆盘型和抛物线型两种不同的结构元素可以取得不错的去噪效果<sup>[29]</sup>。而本文在进行 MMF 分离低频信号和含噪的高频信号过程中, 在保持其他参数不变的情况下, 仅改变结构元素  $M$  的大小, 以研究不同结构元素大小对分离结果的影响。图 5 所示为使用不同大小(例如 10、50、80 和 120)的结构元素, 从相同的仿真信号中提取有效低频部分。当结构元素  $M$  的大小为 10 时, 噪声尺寸大于 10 的噪声不能分离出来, 如图 5(a)所示, 在高质量信号处, MMF 提取出的低频变化特征与原始信号保持一致, 但是没有达到分离仿真噪声的效果。当结构元素增大到 50 时, 如图 5(b)所示, 可以将噪声尺寸小于 50 的噪声分离出来, 但是依旧没有达到低频信号与含噪高频信号完全分离效果。当结构元素增大到 80 左右时, 如图 5(c)所示, 可以将加入的仿真噪声与低频信号分离出来, 且提取出的低频信号

与原始高质量低频信号保持一致。当结构元素继续增大时, 如图 5(d)所示, 会造成低频信号的提取效果不佳, 甚至可能会造成新的噪声干扰, 因此, 只有对结构元素的大小选取恰当, 才能达到低频和高频分离的效果。

K-SVD 字典学习过程中的阈值  $A$  是分离高频噪声和高频有效信号的一个重要参数, 通过选取合适的阈值才能将噪声的轮廓给提取出来, 从而将高频噪声从高频信号中剔除, 达到信噪分离的目的。选取不同的阈值  $A$  构造的仿真噪声轮廓效果如图 6 所示。图 6(a)中的阈值过小时, 信号过处理, 将原始高质量大地电磁信号作为噪声处理, 从而造成一部分高质量信号的丢失。图 6(b)中的阈值选取适当, 提取的噪声轮廓与加入噪声轮廓基本一致, 将人为加入的噪声全部提取出来了, 达到很好的信噪分离效果。图 6(c) 中的阈值过大时, 对于加入的噪声的没有完全提取出来, 在加入噪声处仍存在一点误差, 导致噪声没有去除干净, 造成新的噪声干扰。通过选择合适的阈值, 将噪声轮廓提取出来, 以达到最佳的信噪分离效果。

对加入噪声的  $E_x$ 、 $E_y$  方向的 MT 数据进行处理, 经过 MMF-KSVD 处理后, 选取  $E_y$  其中一段如图 7 所示。其中图 7(a)所示为加噪后的仿真信号, 图 7(b)所示为经过 MMF 处理后的保留的低频有效信号, 图 7(c) 所示为 MMF 处理后的分离的含噪高

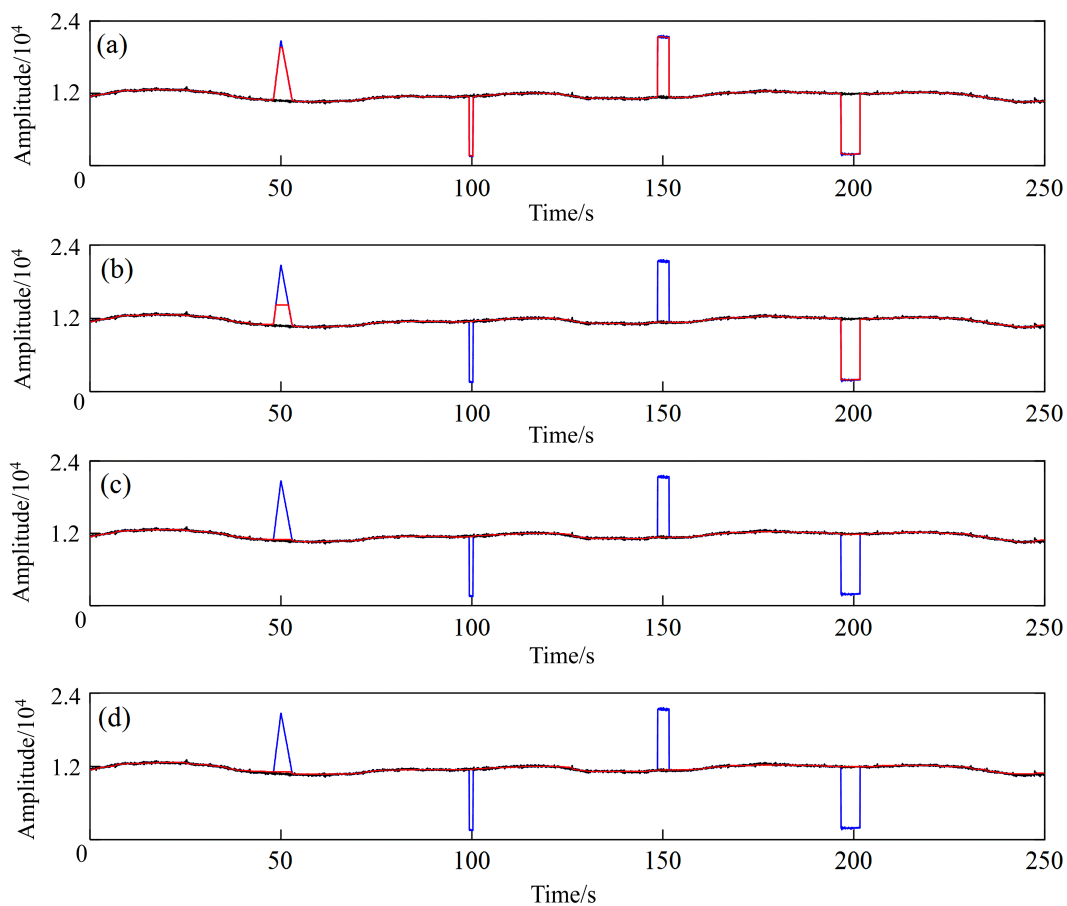


图 5 不同结构元素大小 MMF 分离效果(红: MMF; 蓝: Noisy; 黑: Original)

**Fig. 5** Effect of different structural element sizes of MMF (red: MMF; blue: Noisy; black: Original): (a) MMF-10; (b) MMF-50; (c) MMF-80; (d) MMF-120

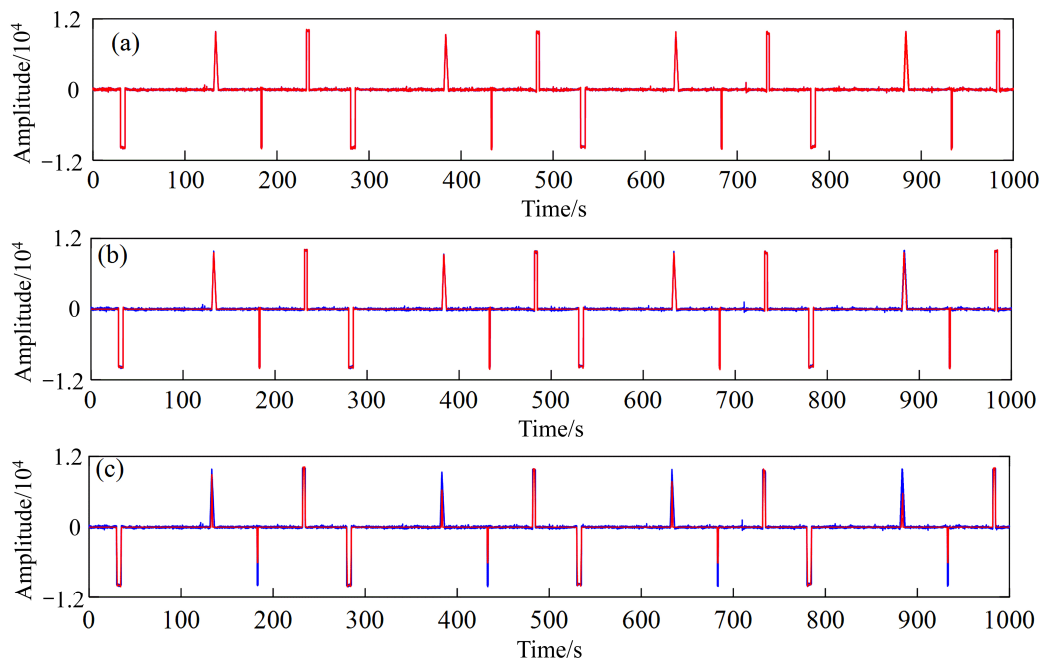


图 6 不同大小的阈值提取的噪声轮廓

**Fig. 6** Noise profiles extracted by selecting different threshold sizes: (a) KSVD-50; (b) KSVD-500; (c) KSVD-8000

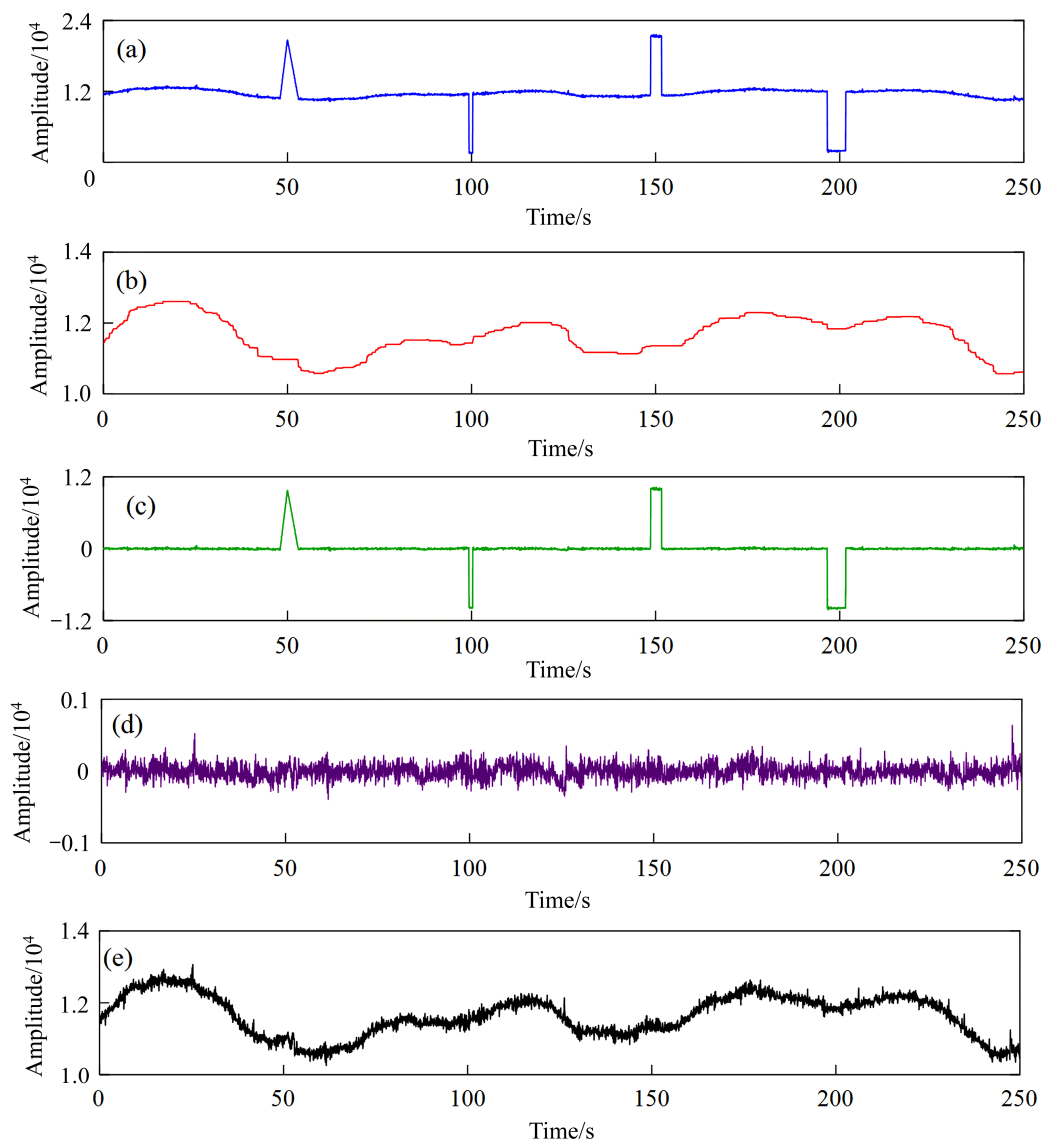


图7 MMF-KSVD 处理仿真噪声信号

Fig. 7 Simulation results of MMF-KSVD: (a) Noisy; (b) MMF-low; (c) MMF-high; (d) High-KSVD; (e) MMF-KSVD

频信号。不难看出, MMF 处理过后, 信号的低频特性被很好保留下来。随后对分离出来的含噪高频信号进行 K-SVD 处理, 处理过后的高频信号几乎不存在仿真噪声的干扰。将 K-SVD 处理后的高频有效信号与图 7(b)中的 MMF 处理后的低频有效信号进行拼接, 得到 MMF-KSVD 处理后的时间序列。经过 MMF-KSVD 处理后, 噪声基本去除干净。将加入仿真噪声的  $E_x$ 、 $E_y$  方向都进行 MMF-KSVD 处理, 以获得全频带 MT 数据。

为了验证 MMF-KSVD 方法的可行性, 本文将 MMF-KSVD 处理后的信号与其他方法(单一的 MMF、K-SVD 和 MMF-Wavelet)的结果进行对比。

图 8(a)所示为原始高质量信号; 图 8(b) 所示为加入噪声的仿真信号; 图 8(c) 所示为只经过单一 MMF 处理后的信号, 与原始高质量信号对比后, 将噪声给分离出来了, 同时丢失了低频信号; 图 8(d) 所示为经过单一 K-SVD 处理后的信号, 也丢失了低频信号; 图 8(e) 所示为 MMF-Wavelet 处理后的效果, 在加入噪声的 50 s、100 s、150 s、200 s 的位置, 经过处理后的信号均有很大的改善, 但是与原始高质量信号对比, 仍存在一定的差异; 图 8(f) 所示为 MMF-KSVD 处理效果图, 与原始高质量信号对比发现, 加入的噪声信号基本去除干净, 在 50 s 和 200 s 处存在一些细微的差距, 但是整体效果最佳。

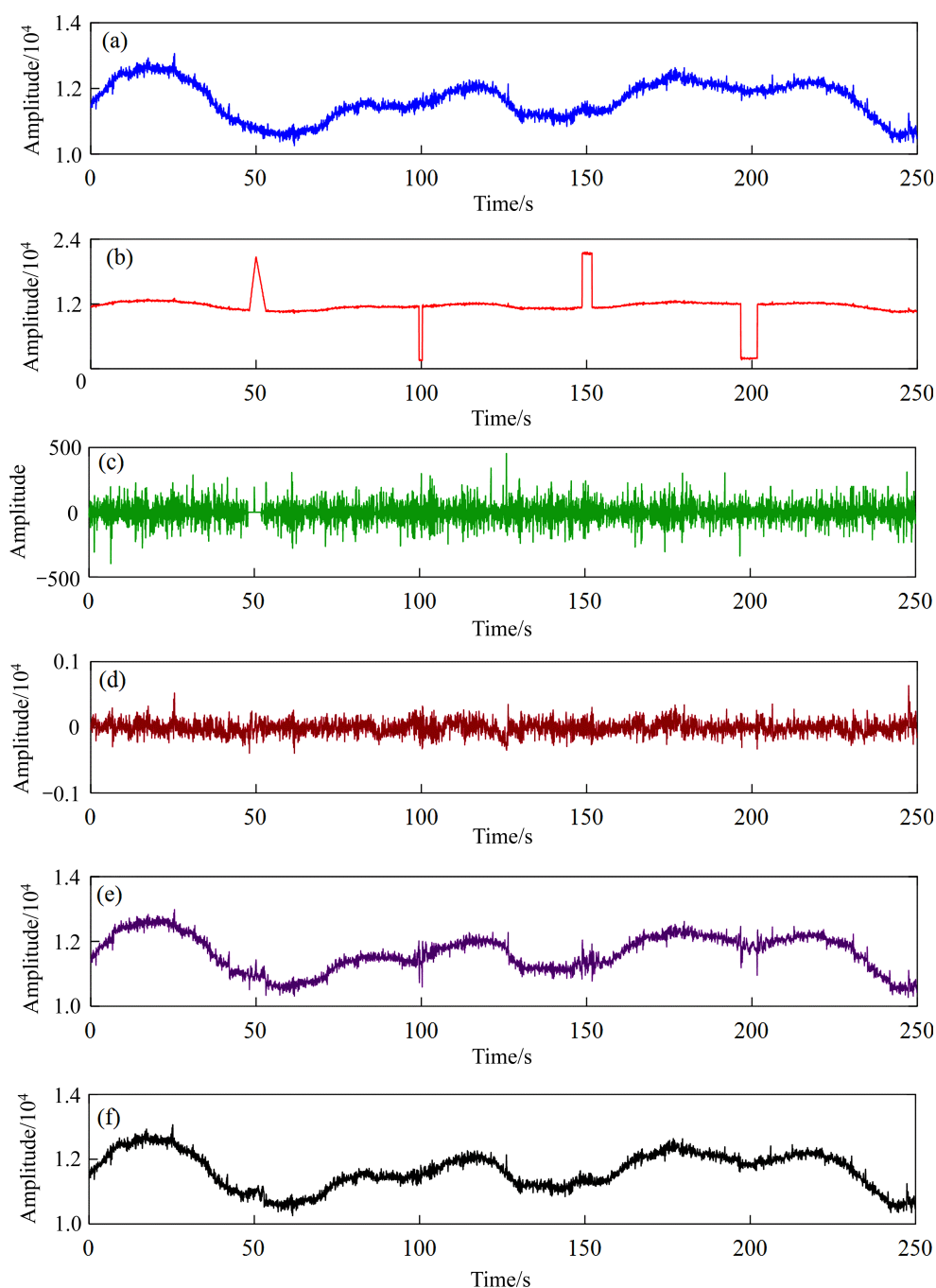


图 8 不同方法的处理时间序列效果对比

**Fig. 8** Comparison of effects of time series use different methods: (a) Original; (b) Noisy; (c) MMF; (d) KSVD; (e) MMF-Wavelet; (f) MMF-KSVD

综上所述, 在没有丢失低频信号也没有损失高频有效信号的情况下, MMF-KSVD 对于仿真信号处理效果最佳。

## 2.2 频谱分析

对比不同方法处理前后低频部分(1 Hz 左右)的

处理效果, 选取 1 Hz 左右的频谱进行处理前后对比。图 9(a)所示为原始高质量信号; 图 9(b)所示为加入仿真噪声后的频谱, 频谱图中幅值大于原始 MT 信号; 图 9(c)所示为经过单一 MMF 处理后的噪声信号的频谱图, 与原始高质量的频谱图对比发现差异较大, 明显丢失了低频信号; 图 9(d)所示为

单一 K-SVD 字典学习处理效果, 信号在低频部分的频谱与原始高质量差别很大, 也丢失了低频信号, 单一的 MMF 或 K-SVD 字典学习容易丢失低频信号; 图 9(e)所示为 MMF-Wavelet 处理后的信号, 信号的频谱图与原始高质量信号基本无差别; 图 9(f)所示为 MMF-KSVD 处理后的频谱, 与原始高质量的频谱基本一致。对比发现, 经过 MMF-KSVD 和 MMF-Wavelet 的处理后, 在最大限度保留低频信号的前提下, 不会造成信号的丢失。

### 2.3 定量分析

仿真实验中采用曲线相似度(Normalized cross correlation, NCC,  $C_{NC}$ )、重构误差(Reconstruction error, RE,  $E_R$ )和信噪比(Signal to noise ratio, SNR,  $R_{SN}$ )三个参数对两种组合方法的处理结果进行定量评价。

重构误差  $E_R$  定义为:

$$E_R = \frac{\|y(x) - r(x)\|_2}{\|y(x)\|_2} \quad (15)$$

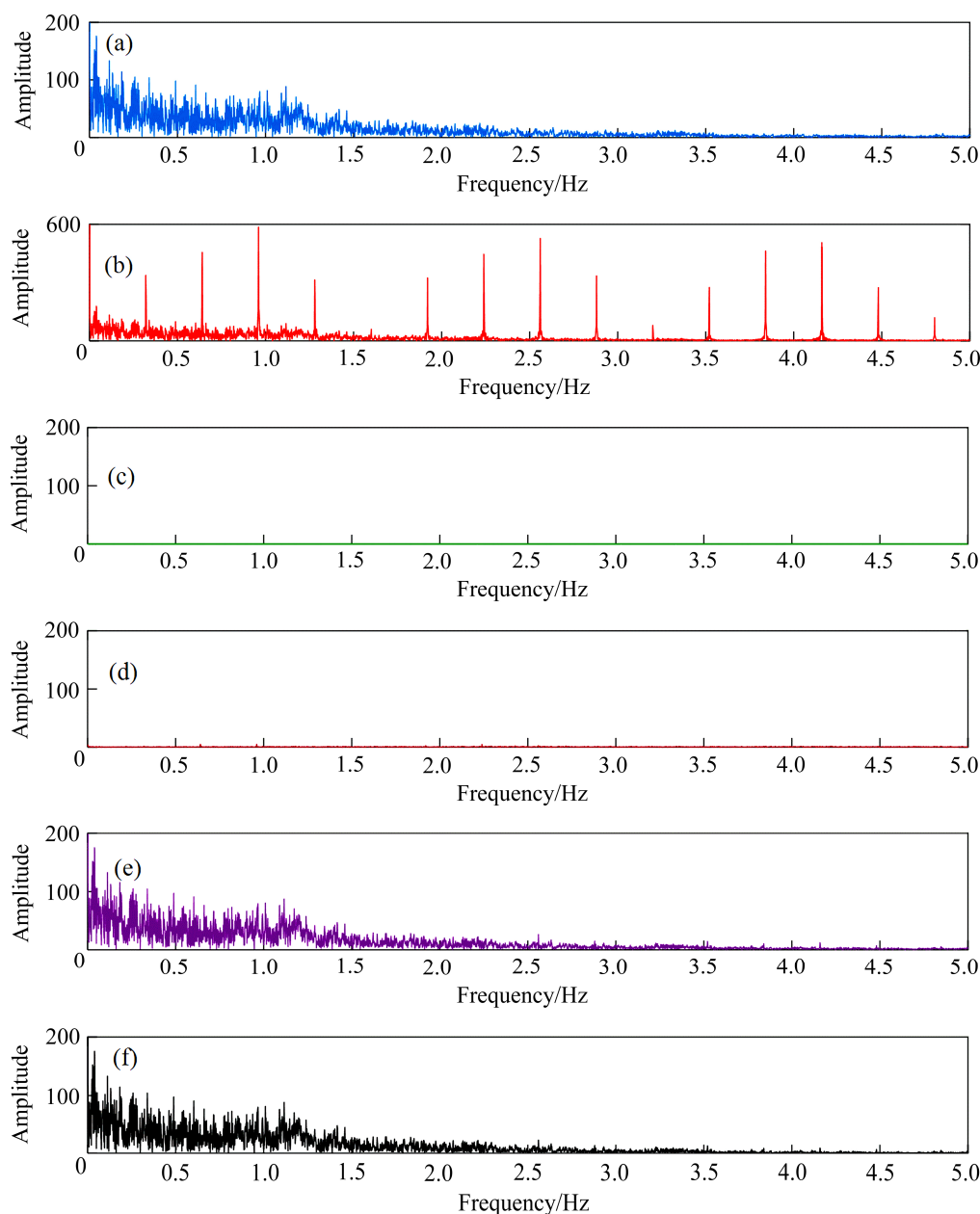


图9 不同方法处理后的信号频谱图对比

**Fig. 9** Spectrum comparison of different methods: (a) Original; (b) Noisy; (c) MMF; (d) KSVD; (e) MMF-Wavelet; (f) MMF-KSVD

曲线相似度  $C_{NC}$  定义为:

$$C_{NC} = \frac{\sum_{x=1}^N y(x) \cdot r(x)}{\sqrt{\left(\sum_{x=1}^N y^2(x)\right) \cdot \left(\sum_{x=1}^N r^2(x)\right)}} \quad (16)$$

信噪比  $R_{SN}$  定义为:

$$R_{SN} = 20 \lg \frac{\|y(x)\|_2}{\|y(x) - r(x)\|_2} \quad (17)$$

式中:  $y(x)$  表示原始高质量信号;  $r(x)$  表示重构信号;  $x$  为采样点的序号;  $N$  为信号长度。

重构误差  $E_R$  为 0 时, 表示重构信号与原始信号完全一样, 因此  $E_R$  越接近 0 表示强干扰分离效果越好。曲线相似度  $C_{NC}$  值为 1 时, 表示两个信号完全一样; 当  $C_{NC}$  为 0 时, 表示两个信号正交; 当  $C_{NC}$  为 -1 时, 表示两个信号绝对值相同而符号相反。因此,  $C_{NC}$  值越接近 1 表示处理过后的信号与原始高质量信号拟合越好。信噪比  $R_{SN}$  越大表示信号的质量越高, 从表 1 中可看出, 两种组合方法对于信号的质量都有很大的改善, 对比  $C_{NC}$ 、 $E_R$  两项参数, 本文方法处理效果明显优于 MMF-Wavelet, 信噪比也从原来的 -5.4166 dB 提高到 20.4848 dB。对比表 1 的各项参数, MMF-KSVD 处理后的仿真信号的各项

表 1 不同方法处理后信号参数变化

Table 1 Changes of various parameters of different methods

| Method      | $C_{NC}$ | $E_R$  | $R_{SN}$ |
|-------------|----------|--------|----------|
| Noisy       | 0.4805   | 1.8656 | -5.4166  |
| MMF-Wavelet | 0.9907   | 0.1361 | 17.3200  |
| MMF-KSVD    | 0.9955   | 0.0946 | 20.4848  |

项参数均优于 MMF-Wavelet, 且效果更佳。

## 2.4 视电阻率-相位分析

通过计算实测点 QH301603 相关的视电阻率-相位来判断 MT 信号处理效果, 从图 10 所示  $xy$  和  $yx$  方向的视电阻率-相位对比发现, 在加入噪声后的视电阻率-相位都从原始的平滑、有序变成毫无规律, 经过 MMF-Wavelet 处理后的视电阻率-相位都有一定的改善, 但是对比原始高质量的视电阻率-相位处理的效果不是特别理想, 尤其在 0.01~1 Hz 左右视电阻率-相位。而 MMF-KSVD 处理后的视电阻率-相位在  $xy$  方向和  $yx$  方向变化趋势与原始高质量的视电阻率-相位基本一致, 说明本文的方法将人为加入的噪声都处理干净, 也没有造成新的噪声干扰, 在一定的程度上说明了 MMF-KSVD 的可行性, 并且处理效果比 MMF-Wavelet 更佳。

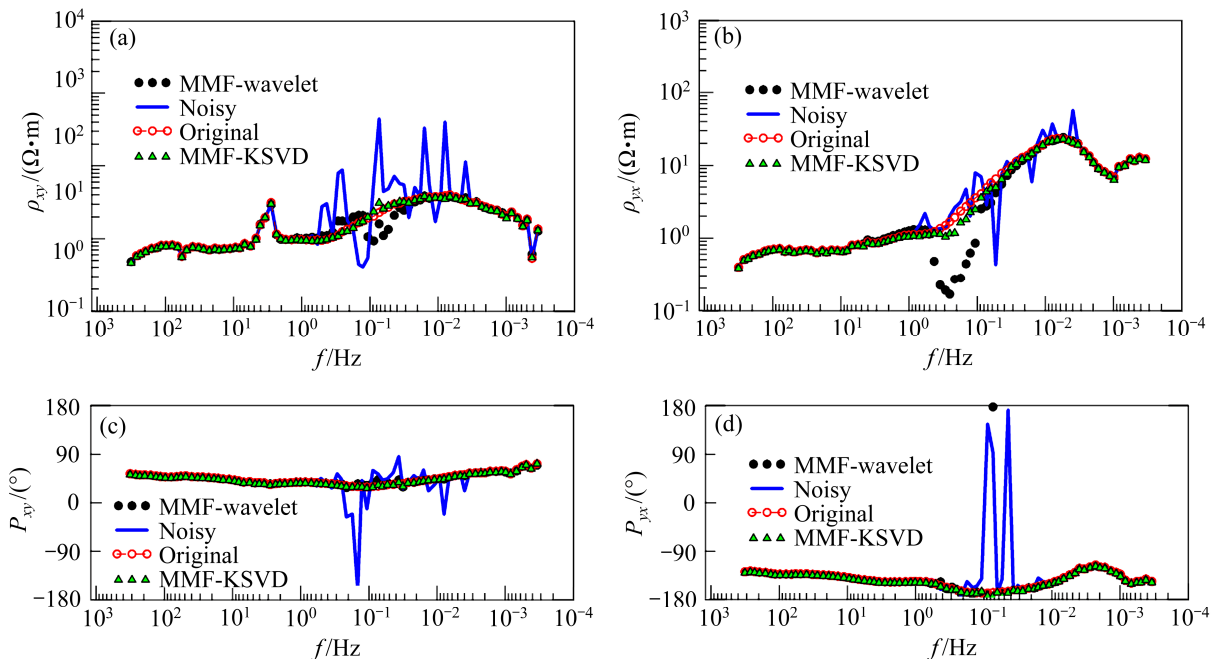


图 10 处理前后视电阻率-相位变化

Fig. 10 Apparent resistivity-phase changes before and after processing with different methods: (a)  $\rho_{xy}$ ; (b)  $\rho_{yx}$ ; (c)  $P_{xy}$ ; (d)  $P_{yx}$

### 3 实测数据处理

#### 3.1 时间序列处理

选取实测数据中存在不同类型噪声信号的时间段,经过MMF-KSVD处理后,成功将噪声分离出来,在无噪声干扰的地方,保持原样,只针对存在噪声的时间序列进行处理,且处理效果明显,将大幅度的噪声从高质量的MT信号中剔除。如图11所示实测数据中存在不同脉宽的方波干扰时,经过MMF-KSVD处理过后,成功提取出方波的噪声轮廓,分离出方波噪声,提高数据质量。如图12所

示,实测数据中存在不同幅值的脉冲噪声干扰时,经过处理后也将噪声分离出来了。如图13所示,在实测点QH101503的 $E_x$ 、 $E_y$ 、 $H_x$ 、 $H_y$ 的同一个时间段存在不同幅值和脉宽的三角噪声干扰,经过MMF-KSVD处理后,成功将噪声剔除,信号的质量得到很大改善。

#### 3.2 视电阻率-相位分析

选取两个点的实测数据,对比处理前后的视电阻率-相位的变化。图14、15所示为原始视电阻率-相位与处理后的视电阻率-相位对比,每个点的视电阻率-相位经过处理后都有一个很好的改善,

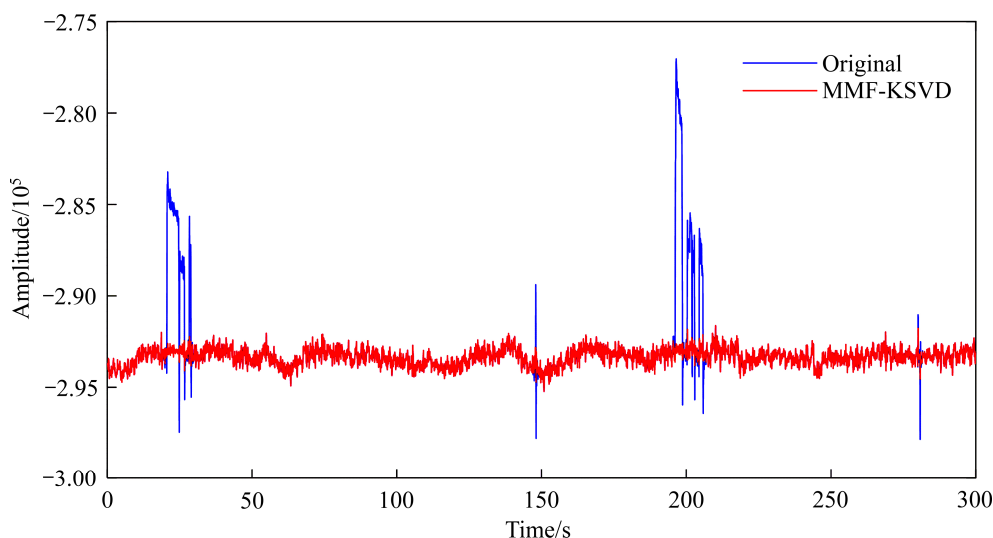


图 11 不同方波噪声分离效果

Fig. 11 Effect of separating different square wave noise

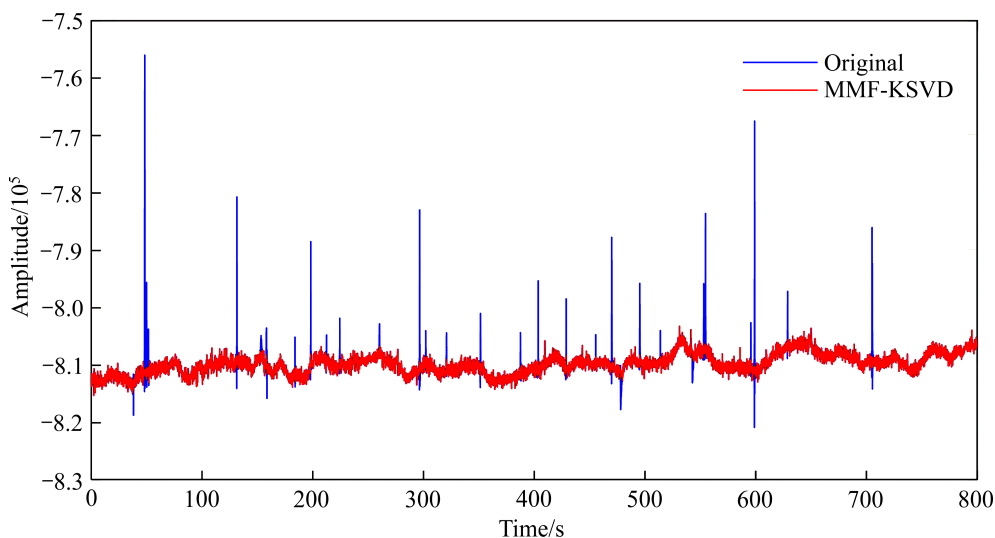


图 12 不同脉冲噪声分离效果

Fig.12 Effect of separating different impulsive noise

在 1 Hz 以下的处理效果明显。实测点 QH101503 视电阻率-相位在  $xy$ 、 $yx$  方向均有很大改善, 视电阻率和相位都从原始的杂乱无章变为光滑、平整、有序; 在 0.01~0.1 Hz 之间存在两个飞点, 后期可

以通过功率谱挑选剔除异常值。实测点 QH301103  $xy$  方向的视电阻率-相位有很大改善, 处理过程中  $E_x$ 、 $E_y$ 、 $H_x$ 、 $H_y$  都进行了 MMF-KSVD 去噪处理, 在噪声较少的  $yx$  方向经过 MMF-KSVD 处理过后,

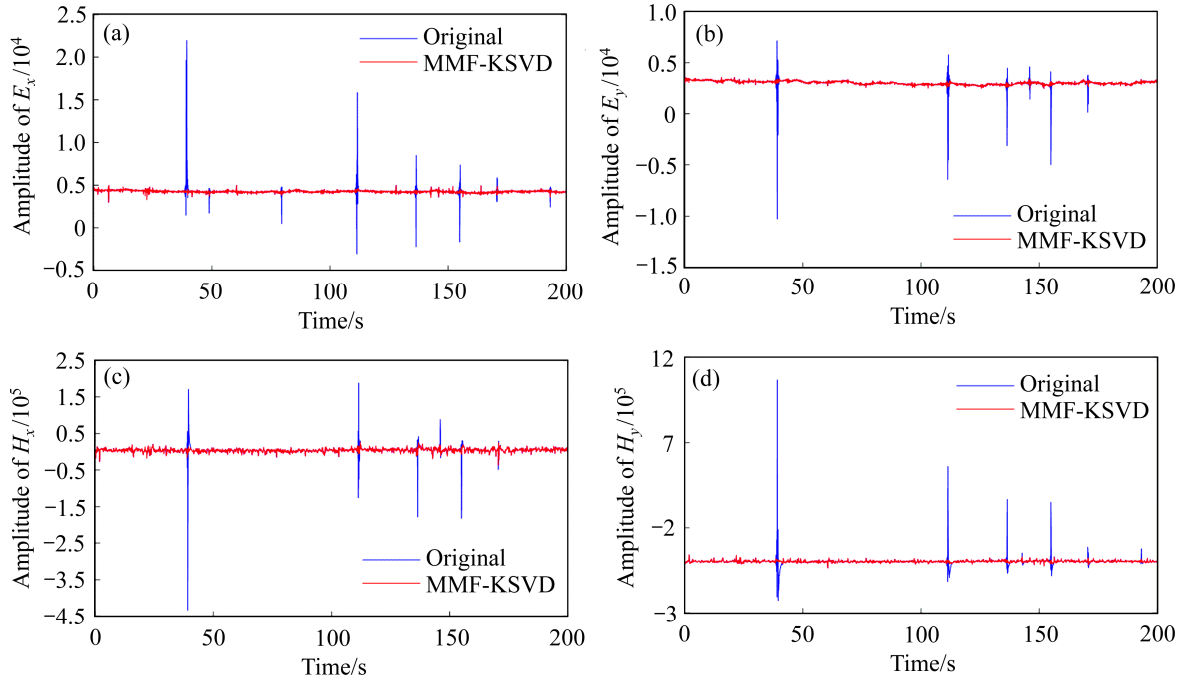


图 13 实测点 QH101503 存在噪声的  $E_x$ 、 $E_y$ 、 $H_x$ 、 $H_y$  方向时间域信噪分离效果

Fig. 13 Time domains of  $E_x$ ,  $E_y$ ,  $H_x$ ,  $H_y$  signal-to-noise separation effect of measured site QH101503 with noise in Qinghai Province, China

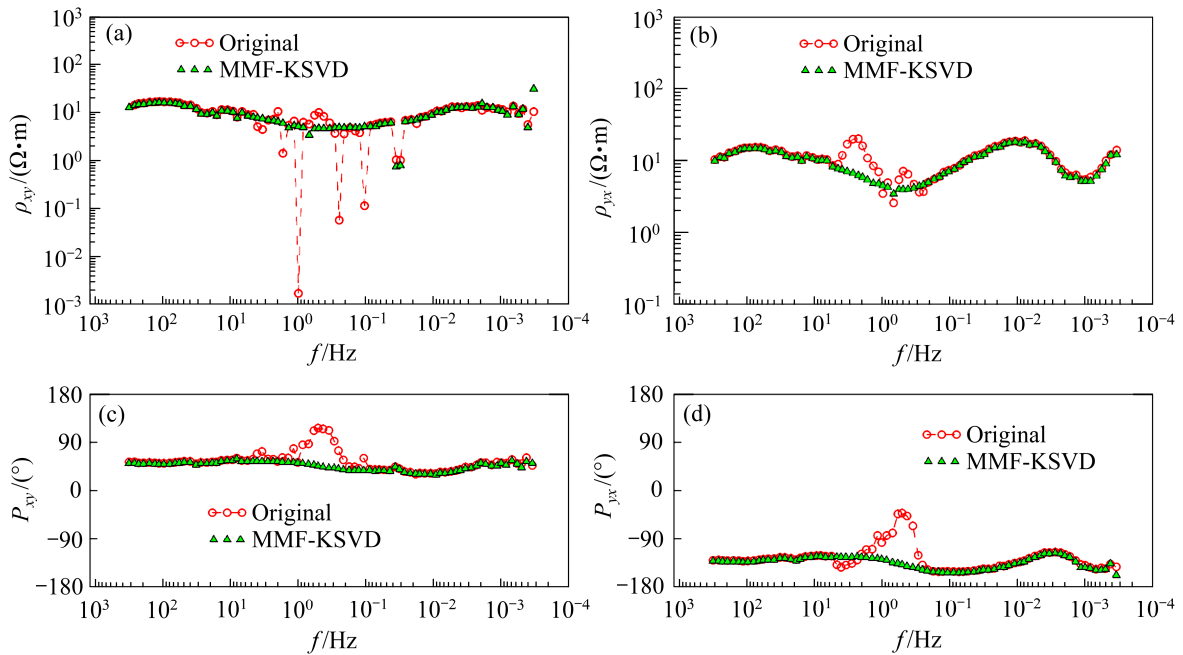


图 14 实测点 QH101503 处理效果

Fig. 14 Effect of apparent resistivity-phase processing at QH101503 measured site in Qinghai Province, China: (a)  $\rho_{xy}$ ; (b)  $\rho_{yx}$ ; (c)  $P_{xy}$ ; (d)  $P_{yx}$

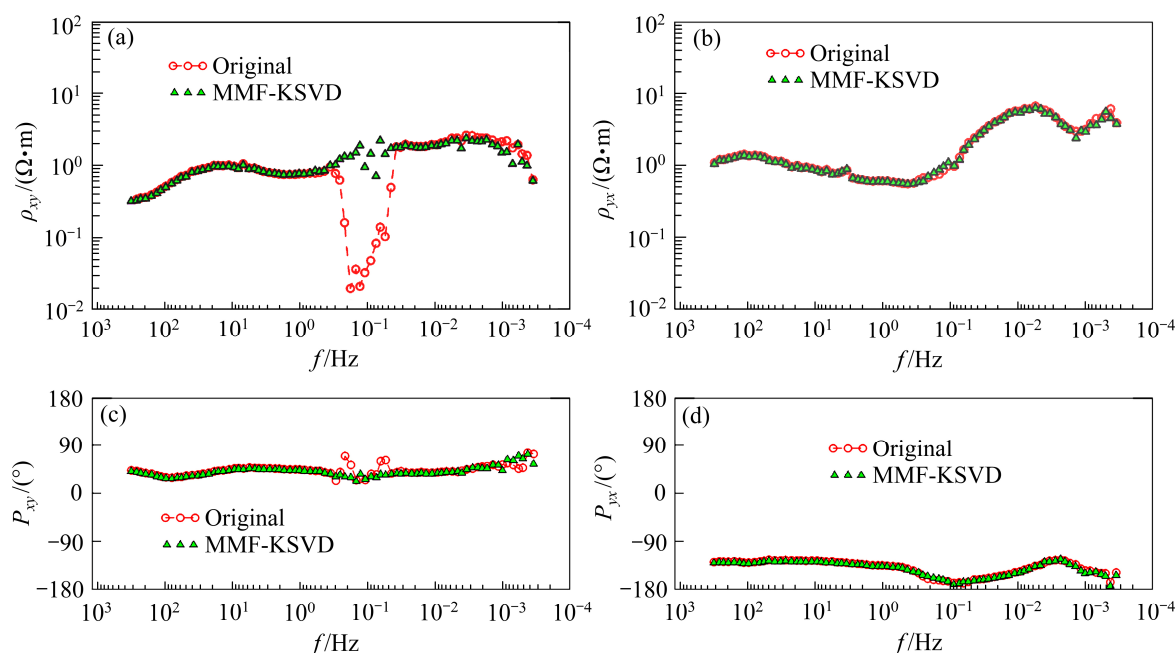


图 15 实测点 QH301103 处理效果

Fig. 15 Effect of apparent resistivity-phase processing at QH301103 measured site in Qinghai Province, China: (a)  $\rho_{xy}$ ; (b)  $\rho_{yx}$ ; (c)  $P_{xy}$ ; (d)  $P_{yx}$

视电阻率-相位处理前后保持一致, 而  $xy$  方向的视电阻率-相位有很大的改善, 说明在处理的过程中只针对噪声信号, 在不造成 MT 信号的丢失的基础上, 提高数据质量。

## 4 结论

针对 1Hz 附近的低频噪声和 MT 信号丢失的问题, 提出了 MMF-KSVD 时间序列编辑方法, 用于提高 MT 数据的信噪比。该方法的优势在于不需要直接从原始数据中提取噪声, 而是先提取有效的低频部分, 对低频信号进行保存, 然后采用 K-SVD 字典学习自主学习到噪声的特征结构, 提取噪声轮廓, 最后实现信噪分离。取得的认识主要有:

1) MMF-KSVD 方法可以分离不同种类的人文噪声并保留低频信号, 效果优于传统方法, 且耗时短, 适用于批量 MT 数据的处理。

2) MMF-KSVD 方法对于降低 1Hz 附近 MT 数据的噪声非常有效, 对于后期的反演和解释具有重要意义。

3) MMF-KSVD 难以改善低于 0.01 Hz 的视电阻率-相位曲线的质量。原因是形态滤波难以准确

的提取宽度大于结构元素长度的方波噪声。这类噪声具有极低的频率, 通常会导致 0.01 Hz 以下的测深曲线失真。对于 MT 信号中超低频人文噪声的压制需要进一步研究。

## REFERENCES

- [1] DENG Ju-zhi, YU Hui, CHEN Hui, et al. Ore-controlling structures of the Xiangshan volcanic Basin, SE China: Revealed from three-dimensional inversion of Magnetotelluric data[J]. Ore Geology Reviews, 2020, 127: 103807.
- [2] 邓居智, 陈辉, 殷长春, 等. 九瑞矿集区三维电性结构研究及找矿意义[J]. 地球物理学报, 2015, 58(12): 4465-4477, Doi: 10.6038/cjg20151211.  
DENG Ju-zhi, CHEN Hui, YIN Chang-chun, et al. Three-dimensional electrical structures and significance for mineral exploration in the Jiujiang-Ruichang District[J]. Chinese Journal of Geophysics, 2015, 58(12): 4465-4477.
- [3] 李爱勇, 曹创华, 柳建新, 等. 不同幅度噪声和缺失数据对大地电磁正则化反演的影响[J]. 中国有色金属学报, 2012, 22(3): 915-921.  
LI Ai-yong, CAO Chuang-hua, LIU Jian-xin, et al. Effects of

- different amplitudes of noise and missing data on regular inversion of magnetotelluric response[J]. The Chinese Journal of Nonferrous Metals, 2012, 22(3): 915–921.
- [4] 肖 晓, 汤井田, 周 聪, 等. 庐枞矿集区大地电磁探测及电性结构初探[J]. 地质学报, 2011, 85(5): 873–886.  
XIAO Xiao, TANG Jing-tian, ZHOU Cong, et al. Magnetotelluric sounding in the Lujiang-Zongyang ore-district and preliminary study of electrical structure[J]. Acta Geologica Sinica, 2011, 85(5): 873–886.
- [5] GOUBAU W M, GAMBLE T D, CLARKE J. Magnetotelluric data analysis: removal of bias[J]. Geophysics, 1978, 43(6): 1157–1166.
- [6] EGBERT G D, BOOKER J R. Multivariate analysis of geomagnetic array data: 1. The response space[J]. Journal of Geophysical Research: Solid Earth, 1989, 94(B10): 14227–14247.
- [7] NEUKIRCH M, GARCIA X. Nonstationary magnetotelluric data processing with instantaneous parameter[J]. Journal of Geophysical Research: Solid Earth, 2014, 119(3): 1634–1654.
- [8] 汤井田, 张 弛, 肖 晓, 等. 大地电磁阻抗估计方法对比[J]. 中国有色金属学报, 2013, 23(9): 2351–2358.  
TANG Jing-tian, ZHANG Chi, XIAO Xiao, et al. Comparison of methods for magnetotelluric impedance estimation[J]. The Chinese Journal of Nonferrous Metals, 2013, 23(9): 2351–2358.
- [9] PARKER R L, BOOKER J R. Optimal one-dimensional inversion and bounding of magnetotelluric apparent resistivity and phase measurements[J]. Physics of the Earth and Planetary Interiors, 1996, 98(3/4): 269–282.
- [10] 王 辉, 程久龙, 姚郁松, 等. 基于站间天然电磁场单位脉冲响应的大地电磁时间序列去噪方法[J]. 地球物理学报, 2019, 62(3): 1057–1070.  
WANG Hui, CHENG Jiu-long, YAO Yu-song, et al. A new method of noise deletion in magnetotelluric time-series based on impulse response of inter-station transfer function[J]. Chinese Journal of Geophysics, 2019, 62(3): 1057–1070.
- [11] 凌振宝, 王沛元, 万云霞, 等. 强人文干扰环境的电磁数据小波去噪方法研究[J]. 地球物理学报, 2016, 59(9): 3436–3447.  
LING Zhen-bao, WANG Pei-yuan, WAN Yun-xia, et al. A combined wavelet transform algorithm used for de-noising magnetotellurics data in the strong human noise[J]. Chinese Journal of Geophysics, 2016, 59(9): 3436–3447.
- [12] 范翠松, 李桐林, 王大勇. 小波变换对 MT 数据中方波噪声的处理[J]. 吉林大学学报(地球科学版), 2008, 38(S1): 61–63.  
FAN Cui-song, LI Tong-lin, WANG Da-yong. Treatment of wavelet transform for square wave noise in MT data[J]. Journal of Jilin University (Earth Science Edition), 2008, 38(S1): 61–63.
- [13] 曹小玲, 刘开元, 严良俊. 大地电磁的小波变换-独立分量分析去噪[J]. 石油地球物理勘探, 2018, 53(1): 206–213, 11.  
CAO Xiao-ling, LIU Kai-yuan, YAN Liang-jun. Magnetotelluric data de-noising based on wavelet transform and independent component analysis[J]. Oil Geophysical Prospecting, 2018, 53(1): 206–213, 11.
- [14] 徐义贤, 王家映. 基于连续小波变换的大地电磁信号谱估计方法[J]. 地球物理学报, 2000, 43(5): 677–683.  
XU Yi-xian, WANG Jia-ying. Power spectrum estimation for magnetotelluric signal based on continuous wavelet transform[J]. Chinese Journal of Geophysics, 2000, 43(5): 677–683.
- [15] 许滔滔, 王中兴, 肖卓伟, 等. 基于 LSTM 循环神经网络的大地电磁工频干扰压制[J]. 地球物理学进展, 2020, 35(5): 2016–2022.  
XU Tao-tao, WANG Zhong-xing, XIAO Zhuo-wei, et al. Magnetotelluric power frequency interference suppression based on LSTM recurrent neural network[J]. Progress in Geophysics, 2020, 35(5): 2016–2022.
- [16] LING Z B, WANG P Y, WAN Y X, et al. A combined wavelet transform algorithm used for de-noising magnetotellurics data in the strong human noise[J]. Chinese Journal of Geophysics, 2016, 59(9): 3436–3447.
- [17] TANG J T, LI G, ZHOU C, et al. Power-line interference suppression of MT data based on frequency domain sparse decomposition[J]. Journal of Central South University, 2018, 25(9): 2150–2163.
- [18] 汤井田, 李 广, 肖 晓, 等. 基于压缩感知重构算法的大地电磁强干扰分离[J]. 地球物理学报, 2017, 60(9): 3642–3654.  
TANG Jing-tian, LI Guang, XIAO Xiao, et al. Strong noise separation for magnetotelluric data based on a signal reconstruction algorithm of compressive sensing[J]. Chinese

- Journal of Geophysics, 2017, 60(9): 3642–3654.
- [19] LI J, ZHANG X, TANG J T, et al. Audio magnetotelluric signal-noise identification and separation based on multifractal spectrum and matching pursuit[J]. Fractals, 2019, 27(1): 1940007.
- [20] 周 聪, 汤井田, 原 源, 等. 大地电磁多参考站阵列数据处理方法[J]. 石油地球物理勘探, 2020, 55(6): 1373–1382, 1167.
- ZHOU Cong, TANG Jing-tian, YUAN Yuan, et al. Multi-reference array MT data processing method[J]. Oil Geophysical Prospecting, 2020, 55(6): 1373–1382, 1167.
- [21] HUANG W, WANG R, ZU S, et al. Low-frequency noise attenuation in seismic and microseismic data using mathematical morphological filtering[J]. Geophysical Journal International, 2020, 222(3): 1728–1749.
- [22] 李 晋, 汤井田, 肖 晓, 等. 基于组合广义形态滤波的大地电磁资料处理[J]. 中南大学学报(自然科学版), 2014, 45(1): 173–185.
- LI Jin, TANG Jing-tian, XIAO Xiao, et al. Magnetotelluric data processing based on combined generalized morphological filter[J]. Journal of Central South University (Science and Technology), 2014, 45(1): 173–185.
- [23] 汤井田, 李 晋, 肖 晓, 等. 数学形态滤波与大地电磁噪声压制[J]. 地球物理学报, 2012, 55(5): 1784–1793.
- TANG Jin-tian, LI Jin, XIAO Xiao, et al. Mathematical morphology filtering and noise suppression of magnetotelluric sounding data[J]. Chinese Journal of Geophysics, 2012, 55(5): 1784–1793.
- [24] OLSHAUSEN B A, FIELD D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images[J]. Nature, 1996, 381(6583): 607–609.
- [25] ENGAN K, AASE S O, HUSOY J H. Method of optimal directions for frame design[C]//Proceedings ICASSP99 (Cat. No.99CH36258). Phoenix, USA: IEEE, 1999: 2443–2446.
- [26] AHARON M, ELAD M, BRUCKSTEIN A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE Transactions on Signal Processing, 2006, 54(11): 4311–4322.
- [27] 汤井田, 李 广, 周 聪, 等. 基于字典学习的音频大地电磁数据处理[J]. 地球物理学报, 2018, 61(9): 3835–3850.
- TANG Jing-tian, LI Guang, ZHOU Cong, et al. Denoising AMT data based on dictionary learning[J]. Chinese Journal of Geophysics, 2018, 61(9): 3835–3850.
- [28] LI G, LIU X Q, TANG J T, et al. Improved shift-invariant sparse coding for noise attenuation of magnetotelluric data[J]. Earth Planets and Space, 2020, 72(1): 4311–4322.
- [29] 汤井田, 李 晋, 肖 晓, 等. 基于数学形态滤波的大地电磁强干扰分离方法[J]. 中南大学学报(自然科学版), 2012, 43(6): 2215–2221.
- TANG Jing-tian, LI Jin, XIAO Xiao, et al. Magnetotelluric sounding data strong interference separation method based on mathematical morphology filtering[J]. Journal of Central South University (Science and Technology), 2012, 43(6): 2215–2221.

## De-noising magnetotelluric data based on mathematical morphology and K-SVD dictionary learning

GUI Tuan-fu<sup>1,2,3</sup>, DENG Ju-zhi<sup>1,3</sup>, LI Guang<sup>1,2,3</sup>, LIU Xiao-qiong<sup>1,2</sup>, CHEN Hui<sup>1,3</sup>, HE Zhu-shi<sup>1,2,3</sup>

(1. Earthquake Prevention and Mitigation and Engineering Geological Disaster Detection Engineering Research Center in Jiangxi Province (East China University of Technology), Nanchang 330013, China;

2. Jiangxi Engineering Laboratory on Radioactive Geoscience and Big Data Technology (East China University of Technology), Nanchang 330013, China;

3. School of Geophysics and Measurement and Control Technology, East China University of Technology, Nanchang 330013, China)

**Abstract:** In order to solve the problem that the conventional magnetotelluric (MT) time-domain denoising methods have the limitation of damaging low-frequency signal, a new method based on mathematical morphological filtering (MMF) and K-singular value decomposition (K-SVD) dictionary learning was proposed to suppress strong cultural noise near 1Hz of MT data. First, MMF was used to separate the low-frequency signal and protect it entirely. Second, K-SVD dictionary learning was used to process residual noisy signals. The field data was used to extract noise contours from auto-learning cultural noise features, which could achieve the purpose of removing noise. The method was verified by testing a synthetic data set and then was used to process two measured data sets. The results show that the proposed method can eliminate all kinds of strong cultural noises without losing useful signals and improve the signal-to-noise ratio and data quality. Moreover, the denoising effect of the proposed method is better than the traditional methods such as wavelet transform.

**Key words:** magnetotelluric; mathematical morphological filtering; K-SVD dictionary learning; strong interference suppression; low-frequency signal

**Foundation item:** Projects(41904076, 41830107, 41674077) supported by the National Natural Science Foundation of China; Project(2018YFC0603202) supported by the National Key Research and Development Program of China; Project(20192BAB212009) supported by the Natural Science Foundation of Jiangxi Province, China; Project(JELRGBDT202003) supported by the Open Fund from Jiangxi Engineering Laboratory on Radioactive Geoscience and Big Data Technology, China; Project(SDGD202008) supported by the Open Fund from Engineering Research Center for Seismic Disaster Prevention and Engineering Geological Disaster Detection of Jiangxi Province, China

**Received date:** 2020-12-17; **Accepted date:** 2021-04-25

**Corresponding author:** DENG Ju-zhi; Tel: +86-13507948149; E-mail: jzdeng73@gmail.com

(编辑 何学锋)