

一种新的基于地震波形指纹的特定地区 核爆炸事件快速检测方法

弓妮 商杰 唐伟 刘哲函 王海军* 黄立洪 韩守诚 江宇

(禁核试北京国家数据中心和放射性核素实验室, 北京 100085)

摘要: 核爆炸监测是禁核试核查的关键技术。为监测全球可能发生的核试验, 全面禁止核试验条约规定了一套严格的核查机制。其中国际监测系统 (International Monitoring System, IMS) 的波形数据实时传输至国际数据中心 (International Data Centre, IDC) 进行处理和分析, 分别在大约 1 小时、4 小时和 6 小时给出三个不同阶段的自动处理结果。对于特定地区的核爆监测, 直接依赖 IDC 结果存在响应滞后和误检率高的问题。为此, 本文提出了一种基于地震波形指纹的快速检测方法 Seisprint。该方法借鉴音频指纹识别思想, 将历史核爆波形作为模板, 利用滑动窗口与特征提取将连续波形压缩为多个二进制指纹, 通过快速相似性匹配与聚类实现核爆事件自动检测并实时报警。采用朝鲜周边两个 IMS 地震台站和我国东北地区四个地震台站记录的朝鲜 6 次地下核试验以及历史天然地震事件数据对 Seisprint 进行测试。Seisprint 生成的指纹可以有效区分核爆与非核爆信号, 且具有较强的抗噪性; 可在数分钟内完成多个地震台站 24 小时连续波形数据的处理, 实现核爆事件的快速准确检出。结果表明, Seisprint 可提高特定地区核爆事件监测的时效性和准确性。

关键词: CTBT; 核爆炸监测; 地震波形指纹; 局部敏感哈希; 自动处理

中图分类号: P315 收稿日期: 2025-07-01

作者简介: 弓妮 (1999-), 女, 研究实习员, 主要研究方向为基于机器学习的地震信号检测. ORCID: 0009-0008-6813-2153. E-mail: gong.ni@ndc.org.cn

*通讯作者: 王海军 (1971-), 男, 博士, E-mail: wang.haijun@ndc.org.cn

A Novel Method of Detecting Underground Nuclear Explosion at Specific Site Based on Seismic Waveform Fingerprint

Gong Ni, Shang Jie, Tang Wei, Liu Zhehan, Wang Haijun^{*}, Huang Lihong, Han Shoucheng, Jiang Yu
(CTBT Beijing National Data Centre and Beijing Radionuclide Laboratory, Beijing 100085, China)

Abstract: (Objective) Nuclear explosion monitoring is a critical technology for nuclear test ban verification. In order to monitor potential nuclear tests worldwide, the Comprehensive Nuclear-Test-Ban Treaty (CTBT) establishes a rigorous verification regime. Within this framework, waveform data from the International Monitoring System (IMS) are transmitted in real time to the International Data Centre (IDC) for processing and analyzing. The IDC delivers automated processing results at three stages: approximately 1 hour, 4 hours, and 6 hours after data acquisition. For region-specific nuclear explosion monitoring, direct reliance on IDC results faces challenges of delayed response and high false detection rates. (Method) To address these limitations, we propose Seisprint, a rapid detection method based on seismic waveform fingerprints. Inspired by audio fingerprinting techniques, Seisprint utilizes historical nuclear explosion waveforms as templates. Continuous seismic waveforms are compressed into multiple binary fingerprints through sliding-window feature extraction. Automated nuclear event detection and real-time alerts are achieved via rapid similarity matching and clustering. The method was tested using data from two IMS seismic stations near North Korea and four seismic stations in northeastern China, covering six historical underground nuclear tests and natural seismic events in North Korea. (Results) The fingerprints generated by Seisprint effectively distinguish nuclear explosions from non-nuclear signals and demonstrate strong robustness against noise. The method processes an entire day of continuous data from multiple seismic stations within just a few minutes, enabling rapid detection of underground nuclear explosion events. (Conclusion) Results confirm that Seisprint promote the timeliness and accuracy of underground nuclear explosion detection in specific area.

Key words: CTBT; Nuclear Explosion Monitoring; Seismic Waveform Fingerprints; Locality-Sensitive Hashing; Automatic Data Processing.

0 引言

1996年9月24日，全面禁止核试验条约(Comprehensive Nuclear-Test-Ban Treaty, CTBT)在纽约联合国总部颁布，CTBT 禁止任何国家或个人进行任何形式的核爆炸试验(Asada, 2002; Bauer and O'Reilly, 2016)。为了监测全球可能发生的违约核试验，CTBT 给出了一套严格的核查机制，其中包括由 337 个核查设施组成的国际监测系统(International Monitoring System, IMS)，以及负责 IMS 数据处理和分析的国际数据中心(International Data Centre , IDC)。IMS 中的地震、水声和次声台站监测数据实时发送到 IDC 进行处理和分析，分别在事件发生后大约 1 小时、4 小时和 6 小时给出三个不同阶段的自动处理公报(SEL1, SEL2, SEL3)。随后，IDC 分析员对自动处理结果 SEL3 进行校正、完善和质量审核，形成审核事件公报(Reviewed Event Bulletin, REB)，上述自动处理结果和人工分析结果作为 IDC 的数据处理产品对授权用户发布(Coyne et al., 2012)。

相对于人工分析结果，IDC 波形数据自动处理系统中事件的误检率大约是 50%，事件漏检率接近 30%(Arora, 2012)。作为全面禁止核试验条约组织重要的技术部门，IDC 一直通过优化系统参数、集成新技术新方法来不断改进数据处理系统的性能。Arora et al.(2009)在科学技术大会 ISS09 期间首次提出 NET-VISA (Network Processing Vertically Integrated Seismic Analysis)，这种方法利用历史事件的特征参数和关联台站的信号检测参数建立概率统计模型，通过贝叶斯理论筛选最佳台站组合，系统地给出了一种新的类似于机器学习的台网事件关联算法。经过十余年的不断完善和测试(Bras et al., 2011; Arora, 2012; Arora et al., 2013; Kushida et al., 2017)，NET-VISA 技术已于 2018 年 1 月正式集成应用在了 IDC 数据处理系统中(Le Bras et al., 2021)。

Ali et al.(2025)的评估结果表明，NET-VISA 的应用使得扫描事件总数增加了 17.90%，平均每日新增 7 个事件；相比于传统依赖于信号特征参数和震相传播走时规律进行聚类的全球关联(Global Association, GA)(Le Bras et al., 1994)算法，NET-VISA 显著降低了漏检率。但在实际应用中，NET-VISA 需要足够多的历史事件和人工分析结果来训练事件和信号特征参数模型。针对特定地区，当可用历史事件数量较少时，NET-VISA 算法性能下降。此外，IDC 自动处理系统主要面向全球地震事件监测，为关联更多台站数据通常需等待约 1 小时才输出检测结果，其时效性难以满足特定地区实时监测任务的需求。鉴于特定地区的地震事件在相同台站记录中往往具有相似波形特征，一些学者提出了基于波形相似性的检测方法，例如波形互相关算法和指纹与相似性阈值检测法(Fingerprint And Similarity Threshold, FAST) (Yoon et al., 2015)等。波形互相关(Cross-Correlation)或模板匹配(Matched Filtering)是地震学中常用的波形检测和相似性分析方法，通过计算归一化相关系数并选择适当阈值，可以在连续波形数据中识别与模板波形相匹配的部分(Senobari et al., 2018)。互相关方法已成功用于朝鲜核试验及大型余震序列的检测(Gibbons and Ringdal, 2012; Bobrov and Kitov, 2011)，提高了震相拾取和事件构建的准确性。当前 IDC 通过预定义高质量的主事件作为波形模板，寻找相似信号，并对检测出的信号进行关联形成事件列表 XSEL，以补充 SEL 中的漏检事件(Bobrov et al., 2014, 2016)。然而，单台站互相关在噪声或复杂环境下误检率较高(Li and Zhan, 2018; Muir et al., 2023)，对长时间的连续数据或海量模板做滑动相关需要巨大的计算量和 I/O，在大规模应用上成为瓶颈(Gibbons and Ringdal, 2006)。

FAST 是斯坦福大学提出的一种基于区域台网的重复地震事件快速检测技术(Yoon et al., 2015; Bergen and Beroza, 2019)，将波形转换为紧凑指纹，通过哈希索引技术快速在数据库中进行相似性搜索，存储和检索都不需加载原始波形，相较于互相关算法，内存占用和 I/O 成

本显著降低，且计算复杂度为近线性复杂度。FAST 凭借计算效率高，可扩展性强的优势，在连续数据处理时能够自动检测到重复出现的地震事件，适用于实时地震监测。Rong et al.(2018)将 FAST 检测管道应用到加州地区，显著提高了重复事件的检出效率，尤其能发现大量之前未记录的微小地震。然而，FAST 算法只能检测到特定时间段内重复发生的事件，而无法检测到仅发生一次的新事件。核爆事件往往多年一次，在长时间窗内缺乏同源模板，使用 FAST 算法存在漏检风险。

为此，本文提出了一种新的基于地震波形指纹的快速检测方法 Seisprint (Seismic Waveform Fingerprint)，将其应用于特定地区核爆炸事件监测中。该方法在 FAST 算法的基础上，借鉴通信领域中的帧同步理论(Shaheen et al., 2015; Sugita and Ito, 2015)，将历史核爆事件的波形作为帧头，通过时域拼接嵌入待处理的连续地震记录数据流前端。在通信系统中，帧头用于标记数据帧的起始并携带识别信息；在 Seisprint 方法中，帧头可作为核爆事件标识，只有当待处理的连续地震记录中出现与该帧头波形相匹配的信号（即帧头时间戳重现）时，才会被初步判定为候选信号。为了提高事件检测的准确性，实际应用时采取多台站联合判据，只有当多个台站记录中同时检测到与帧头相匹配的波形时才形成候选事件。Seisprint 将 FAST 与模板匹配相结合，是一种模板驱动的波形指纹相似性检测法，无需像 NET-VISA 那样依赖大规模历史事件库预训练，只需要少量模板即可实现快速检测。本文利用朝鲜核试验场周边的两个 IMS 地震台站（韩国 KSRS 台阵的子台 KS31、俄罗斯 USRK 台阵的子台 USA0B）和我国东北地区的 4 个国内地震台站（MDJ、CN2、KDN、YNB）记录到的朝鲜 6 次核试验以及发生在试验场周边的历史地震数据，对 Seisprint 方法进行测试评估，验证其在特定地区核爆事件检测中的有效性。

1 地震数据集

本文采用的朝鲜地区地震数据集来源于朝鲜核试验场区周边六个不同方位的地震台站 (CN2、KDN、YNB、MDJ、USA0B、KS31) 的观测记录。数据集包含两部分：(1) 离线事件库，用于评估指纹的鉴别性与抗噪性；(2) 连续波形数据，用于检验 Seisprint 算法在实际流式数据中的端到端检测效果。

在构建离线事件库时，本文选取了朝鲜 2006-2017 年间进行的六次地下核试验事件（用 NKT1-NKT6 表示）、2018–2024 年发生在朝鲜试验场周边 (40° – 42° N, 128° – 130° E) 的 134 个天然地震事件（源自 CTBTO REB 目录，<https://swp.ctbto.org/web/swp/reb>，震级分布见图 1，位置分布见图 2），以及经人工校验的 150 段背景噪声记录，构建出包含核爆、天然地震和背景噪声的综合事件库。所有事件波形均为台站的垂直向通道记录 (Z 通道)，截取自 P 波初至前 10 秒至后 140 秒的时窗（总长 150 秒）；背景噪声数据则从无事件时段随机截取等长片段。

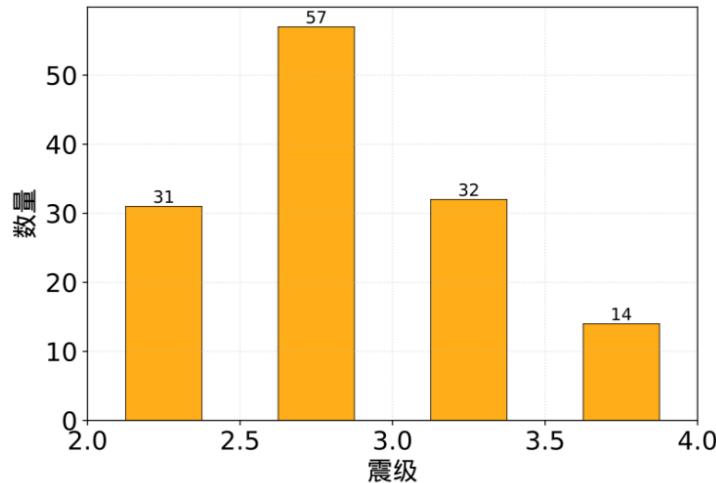


图 1 2018-2024 年间朝鲜核试验场周边天然地震事件震级分布

Fig.1 Magnitude distribution of earthquakes around the North Korean nuclear test site from 2018 to 2024

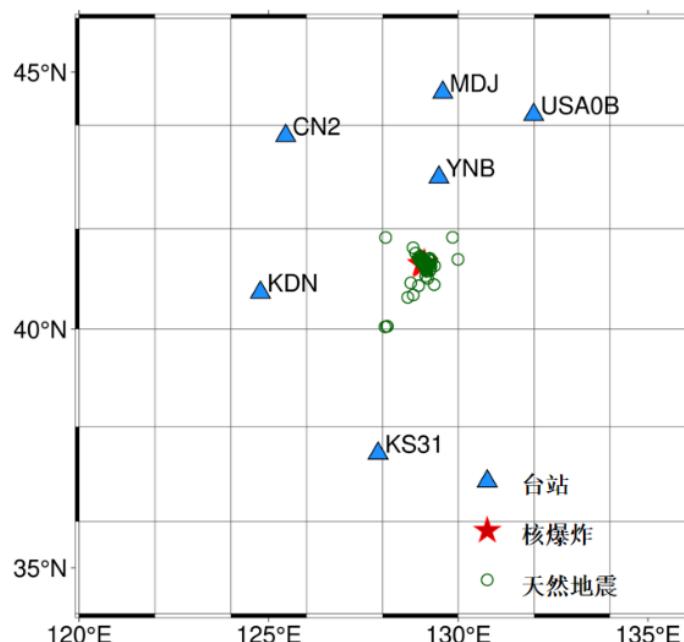


图 2 朝鲜核试验场周边台站位置与事件分布

Fig.2 Station locations and event distribution around the North Korean nuclear test site

蓝色三角形表示台站（其中 KDN、CN2、MDJ、YNB 为我国东北地区的 4 个国内地震台站，KS31 为韩国 KSRS 台阵的子台，USA0B 为俄罗斯 USRK 台阵的子台），红色五角星表示六次核试验 NKT1-NKT6，绿色圆圈表示天然地震事件

连续波形数据部分的选取了共 24 天连续波形，包含以下时间段：（1）NKT3、NKT5、NKT6 三次核试验当天的 24 小时连续波形记录，用于验证对已知核爆事件的检测能力；（2）2016 年 9 月 1 日-8 日、9 月 10 日-20 日共 19 天的连续波形记录，用于评估一般背景下的事件检测能力；（3）NKT6 后发生的三次余震事件（用 PEV1-PEV3 表示，时间分别为 NKT6 事件后约 8 分钟、2017 年 9 月 23 日和 10 月 12 日）当日 24 小时的连续波形记录，用于测试 Seisprint 方法对特定区域内余震事件的检测能力。（需要说明的是，由于 NKT6 与

PEV1 发生在同一天，二者共用同一段连续波形数据）。六次核试验和三次余震事件参数见表 1（来源于 CTBTO REB 目录）。

表 1 朝鲜 2006–2017 年六次核试验（用 NKT1-NKT6 表示）以及 NKT6 后发生的三次余震事件（用 PEV1-PEV3 表示）参数

Table 1 Parameters of the Six Nuclear Test Events in North Korea from 2006 to 2017 (denoted as

NKT1 – NKT6) and three subsequent aftershock events following NKT6 (denoted as PEV1 – PEV3)

朝鲜地区事件	日期	UTC 时间	纬度	经度	震级(mb)
NKT1	2006/10/09	01:35:28	41.29	129.09	4.1
NKT2	2009/05/25	00:54:43	41.31	129.04	4.5
NKT3	2013/02/12	02:57:51	41.30°	129.06°	4.9
NKT4	2016/01/06	01:30:01	41.30°	129.05°	4.8
NKT5	2016/09/09	00:30:01	41.30°	129.05°	5.1
NKT6	2017/09/03	03:30:01	41.32°	129.03°	6.1
PEV1	2017/09/03	03:38:32	41.30	129.07	4.1
PEV2	2017/09/23	08:29:16	41.23	129.35	3.4
PEV3	2017/10/12	16:41:08	41.33	129.17	2.9

2 Seisprint 方法

Seisprint 算法是将 FAST 和模板匹配相结合的一种模板驱动的波形指纹相似性检测方法，处理流程主要包括五个步骤，分别是：预处理、指纹提取、相似性搜索、后处理和事件筛选。各流程相互衔接，协同完成核爆事件的自动检测，其工作流程如图 3 所示。

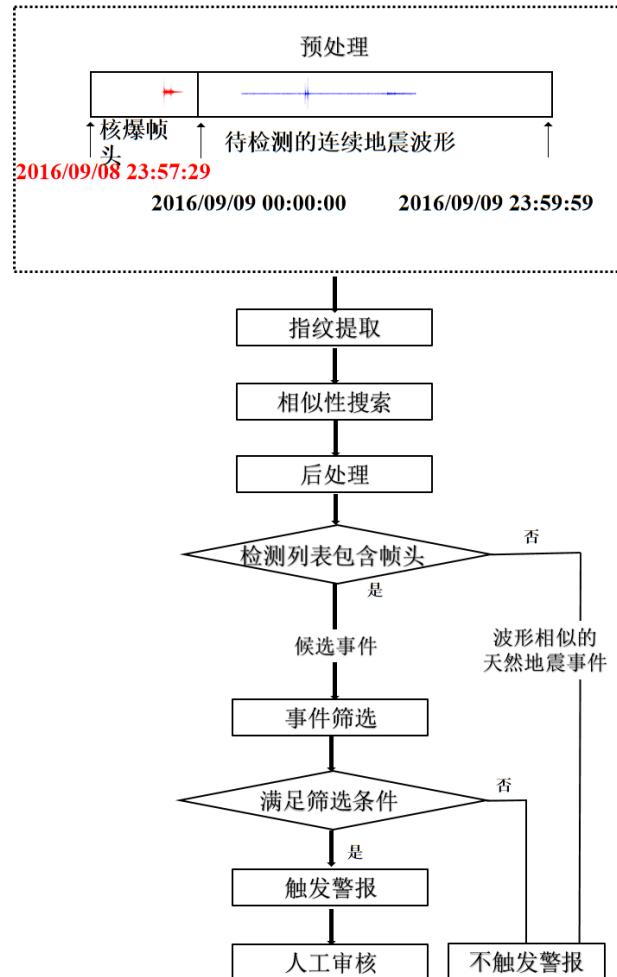


图 3 Seisprint 检测流程
Fig.3 Workflow of the Seisprint Detection Method

2.1 预处理

Seisprint 相较于 FAST 的一项关键改进在于历史核爆波形作为帧头驱动核爆炸事件检测。地下核爆炸在固定场地重复进行时，其震源机制具有高度一致性（各向同性膨胀源），即震源矩张量以各向同性分量为主，对应源区体积的瞬时增加，理论辐射图呈球对称，仅辐射等幅同极性的 P 波而不直接激发 S 波(Richards and Kim , 2005)，假设传播路径的地壳结构相对稳定，爆炸产生的地震波在相同地震台站记录中具有相似性。每次核试验的当量不同(2006 年首次试验当量不足 1 kt，而 2017 年试验推测超过 100 kt) (赵连锋等, 2017)，震源埋深也有所变化，这直接影响地震波的绝对振幅和高频能量含量：较大当量和较深埋深的试验通常产生更强的低频能量，且高频部分受到不同程度的衰减，从而导致同一台站记录的核爆炸波形有所差异(图 4)，但是整体呈现出核爆炸的典型地震记录特征。其相同的波形特征为：

- (1) P 波起始陡峭、Lg 波相对较弱：爆炸源在起爆瞬间释放巨大能量，产生高频冲击波，导致 P 波初至尖锐且振幅陡增；
- (2) P/S 波能量分配异常：P 波高频成分相较于 S 波更丰富，且 P 波幅值大于 S 波幅值；
- (3) 短周期 Rayleigh (Rg) 波发育：Rg 为 Rayleigh 基模在短周期端的导波，常在区域—近区距离内表现为显著能量包。在 MDJ 台站记录的六次核爆事件的波形中，Rg 波主能量稳定落在约 3–5 s 的周期带，显示出典型频散：群速度

随周期变化，与区域尺度的基模 Rayleigh 波频散规律一致(Kafka, 1990)。

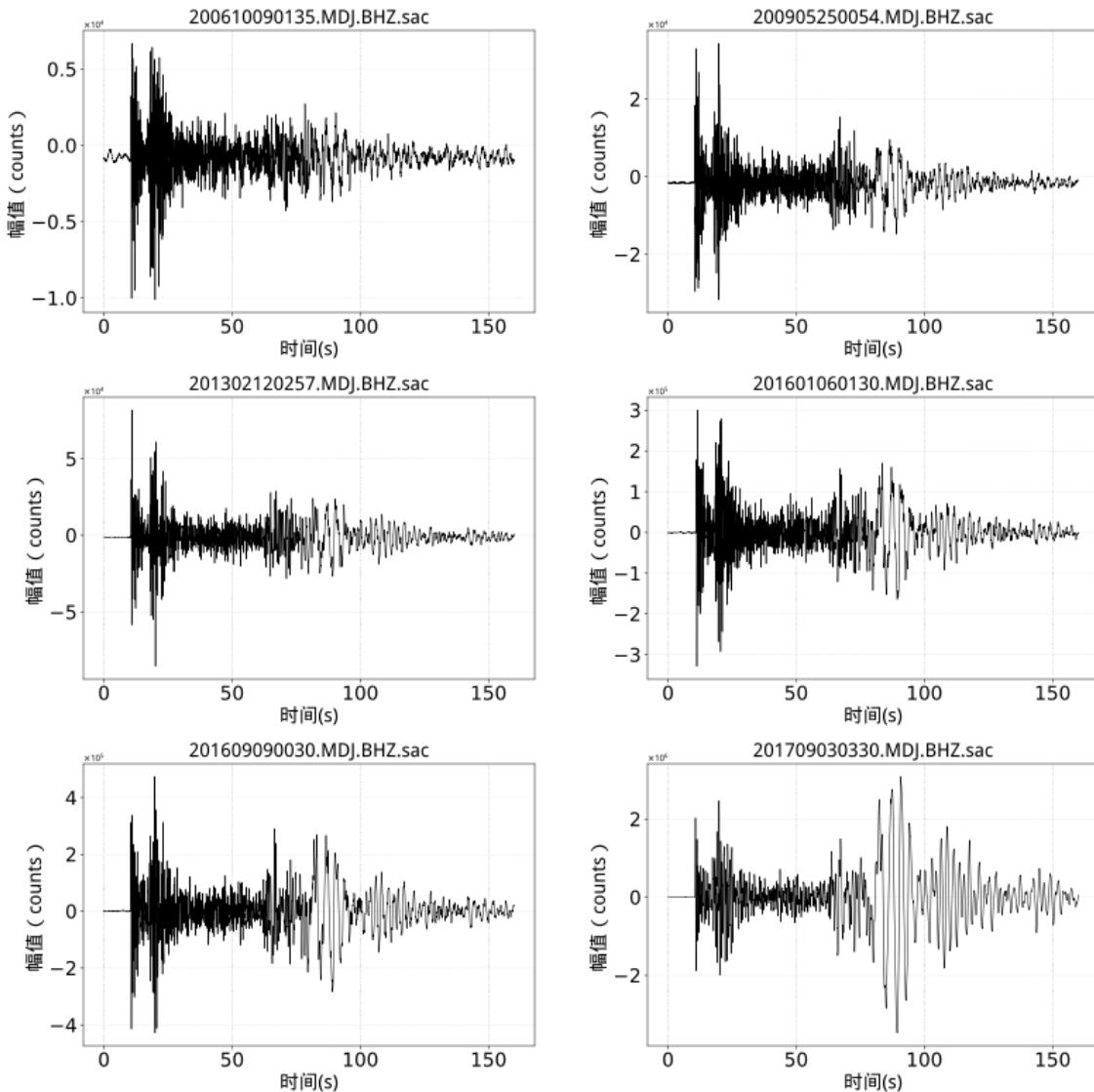


图 4 MDJ 台站记录的六次朝鲜地下核爆事件的波形 (20Hz)

Fig.4 Waveforms of six North Korean underground nuclear test events recorded at MDJ station

Seisprint 算法中，首先对输入的待检测连续波形数据进行预处理：

(1) 模板拼接：选取已知历史核爆事件作为模板，将其在多个台站垂直向通道记录的模板波形窗口 $W_P = [P_{arrival} - T_{before}: P_{arrival} + T_{after}]$ 拼接至待检测的连续波形数据开头。其中， $P_{arrival}$ 为 P 波到时， T_{before} 和 T_{after} 分别为 P 波到时前后截取的窗口长度。

(2) 时间同步：由于添加帧头，需要将待检测波形的起始时间统一向前调整 $T_{before} + T_{after}$ （假设模板波形窗口长度为 150s，待检测数据为 2016 年 9 月 9 日的 24 小时连续波形，那么将拼接后的波形数据起始时间向前调整 150s，设置为 2016 年 9 月 8 日 23:57:30）。

2.2 指纹提取

指纹提取是 Seisprint 检测流程的核心环节，图 5 是 KS31 台记录的 2021 年 6 月 13 日目

标区域内发生的一次天然地震事件与 2009 年 5 月 25 日朝鲜第二次核爆事件（NKT2）的波形（见图 5(a)(b)）及其指纹提取过程。

(1) 首先通过短时傅里叶变换（STFT）将原始地震波形转换为时频谱图，将一维波形映射到“时间 - 频率”平面，捕捉震相能量随时间演化的局部谱结构，计算公式如 (1) 所示：

$$\text{STFT}_x(t, \omega) = \int_{-\infty}^{+\infty} x(\tau)W(\tau - t)e^{-j\omega\tau}d\tau, \quad (1)$$

其中， t 表示分析时刻， ω 为角频率， $x(\tau)$ 表示 τ 时刻的时域地震信号， $W(\tau - t)$ 表示地震信号在 τ 时刻的滑动窗函数， $e^{-j\omega\tau}$ 为傅里叶基函数。STFT 的核心思想是在时间轴上“挪动”一个窗，把以 t 为中心的一段信号拿出来做局部傅里叶变换，因此 $\text{STFT}_x(t, \omega)$ 表示“在时刻 t 附近的频谱”。图 5(c)(d) 展示了天然地震与核爆的时频谱对比，直观反映了二者在不同频段的能量分布差异。

(2) 使用滑动窗口将时频谱图分割生成重叠频谱片段，以增强对时间漂移与小尺度噪声的鲁棒性；然后对每个片段进行二维离散 Haar 小波变换提取特征。Haar 小波变换作为一种对函数进行层次分解的数学工具，能够将图像描述为整体形状与逐级细化细节的叠加，在图像检索领域应用广泛（Stéphane 2009），详细计算过程见附件 1。通过小波变换，把时频图分解为整体轮廓与逐级细节，如图 5(e)(f) 所示。

(3) 使用中位数绝对偏差（MAD）对小波系数进行标准化，使得对异常脉冲与台站噪声更稳健：

$$\begin{aligned} a_j &= \text{median}_i(L_j^{(i)}), \\ b_j &= \text{median}_i(|\text{median}_i(L_j^{(i)} - a_j)|), \\ L'_j^{(i)} &\leftarrow \frac{L_j^{(i)} - a_j}{b_j}, \end{aligned} \quad (2)$$

其中， median 表示计算中位数， $L_j^{(i)}$ 为第 i 个指纹的第 j 个小波系数， $L'_j^{(i)}$ 为标准化后的小波系数； a_j 和 b_j 分别表示第 j 个小波系数在全部指纹集中分布的中心位置和离散度度量。

(4) 为了增强对噪声的鲁棒性，我们对小波图像进行压缩，选取绝对幅值最大的 K 个标准化小波系数，保留系数的符号（正为 +1，负为 -1），其余系数置零。如图 5(g)(h) 所示，其中黑色表示 -1，白色表示 1，灰色表示 0，可以看出，小波图像中能量不突出的部分均变成了灰色。选 Top-K 系数并仅保留符号的目的是为了强制保留“最具判别力”的结构；将高维连续系数压缩为稀疏符号图，显著降低存储与匹配计算量，同时减弱了幅值尺度差异对匹配的影响。

(5) 对压缩后的小波图像进行二进制编码生成指纹向量 (+1 → 01, -1 → 10, 0 → 00)，以支持近线性复杂度的大规模相似度搜索。生成的指纹如图 5(i)(j) 所示，其中黑色表示 0，白色表示 1。

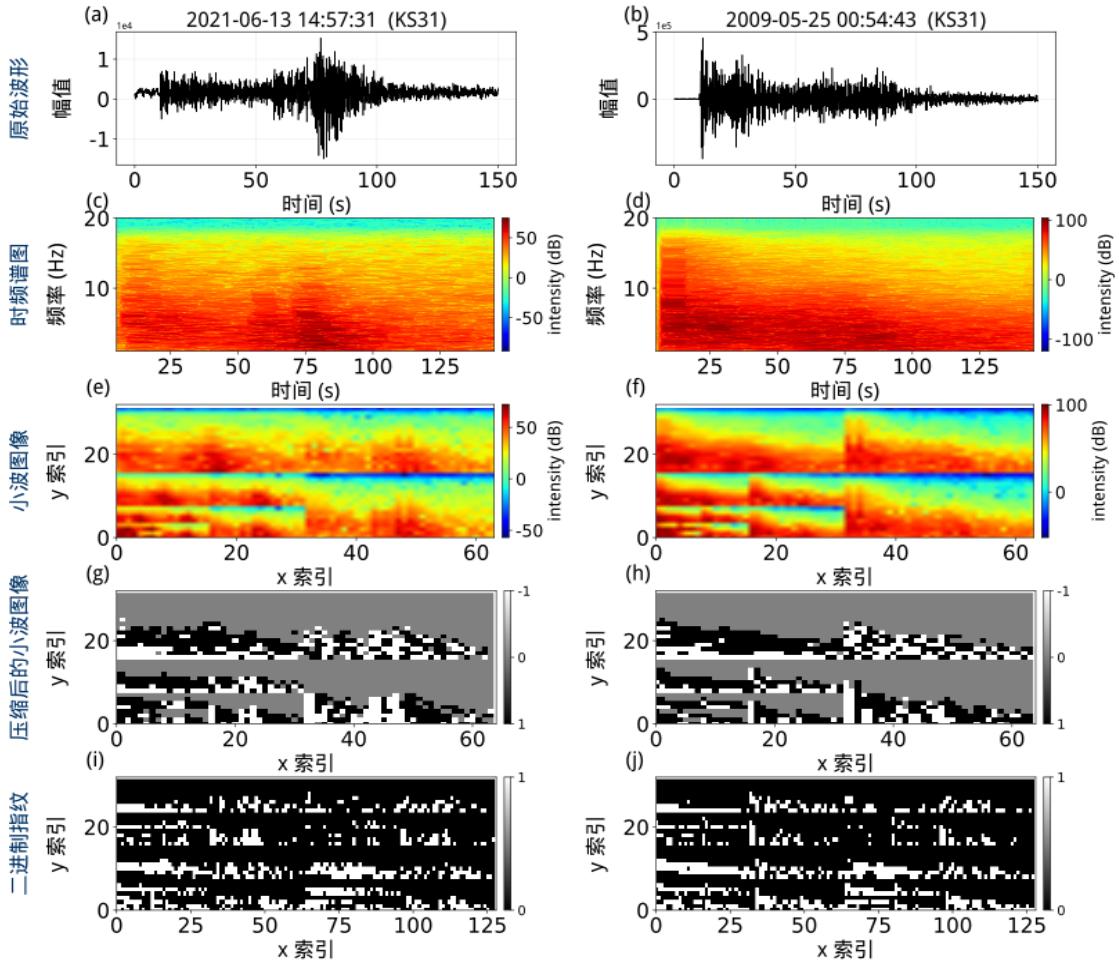


图 5 指纹提取步骤

Fig.5 Fingerprint Extraction Steps

(a)(c)(e)(g)(i) 为 KS31 台站记录到的 2021 年 6 月 13 日发生在目标区域的一次天然地震事件, (b)(d)(f)(h)(j) 为 KS31 台站记录到的朝鲜第二次地下核爆事件; 其波形采样率均为 100Hz;

2.3 相似性搜索

稀疏的二进制指纹可以视为一个集合: 取值为 1 的系数索引构成集合 A, 那么另外一个指纹对应集合 B, 使用 Jaccard 相似度度量两个集合的重叠程度, 定义为:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad (3)$$

基于 Jaccard 相似度, 采用局部敏感哈希(Locality Sensitive Hashing, LSH) (Gionis et al., 1999) 算法实现快速指纹匹配。LSH 是一类为相似度搜索设计的哈希方法: 让“越相似的对象在同一哈希桶中碰撞的概率越高、越不相似的碰撞概率越低”, 从而用少量桶筛选出候选近邻。

本研究中, 指纹提取步骤生成的指纹具有以下特性: 相似波形的指纹相似度显著高于非相似波形的指纹相似度。因此, 采用基于 LSH 的相似性搜索算法, 构成一种基于哈希空间的近似聚类机制, 即相似的波形指纹倾向于多次碰撞, 聚集于相同哈希桶中, 形成稠密的指

纹簇，而不相似的指纹则由于哈希特征离散化，在哈希空间中分布分散，难以跨越相似度阈值，最终被排除在候选列表之外。具体实现分为两步：

(1) Min-Hash 降维：首先利用 Min-Hash 生成哈希签名，Min-Hash 具有重要的 LSH 属性，即两个指纹 fp_1 和 fp_2 的哈希值相等的概率等于其 Jaccard 相似度：

$$\Pr[h(fp_1) = h(fp_2)] = J(fp_1, fp_2), \quad (4)$$

其中， $h(fp_1)$ 表示生成指纹 fp_1 的哈希值。Min-Hash 具体算法细节见(Cohen et al., 2000)，用 p 个独立 Min-Hash 函数 $\mathbf{H} = \{h_1(x), h_2(x), \dots, h_i(x)\}$ 把高维稀疏指纹压缩成长度为 p 的哈希签名 (p 个整数组成的数组)。

(2) LSH 索引与搜索：构建哈希索引库进行近似相似性搜索。将长度为 p 的签名切成 l 个子向量(band)，每个 band 长度为 $m = p/l$ ，把 band “完全相同”的指纹索引映射到相同哈希桶中。如果两个指纹落在同一哈希桶的哈希表个数 $\geq v$ (v 为相似度阈值)，就会判定其相似。匹配搜索时，用相同切分与哈希把查询项落入这 l 个哈希表的若干桶中，并收集桶内候选指纹。

图 6 展示了将波形指纹存储到哈希数据库的过程。哈希数据库包含多个哈希表，每个哈希表包含多个哈希桶。如果两个指纹的 band 相同，就会被哈希到相同哈希桶中，如图中哈希表 2、哈希表 $(l-1)$ 和哈希表 l 所示，否则被哈希到不同的哈希桶中，如哈希表 1 所示。图中两个指纹在 3 个哈希表中被哈希到相同的哈希桶，通过比较 3 与 v 的大小来判断这两个指纹是否相似。

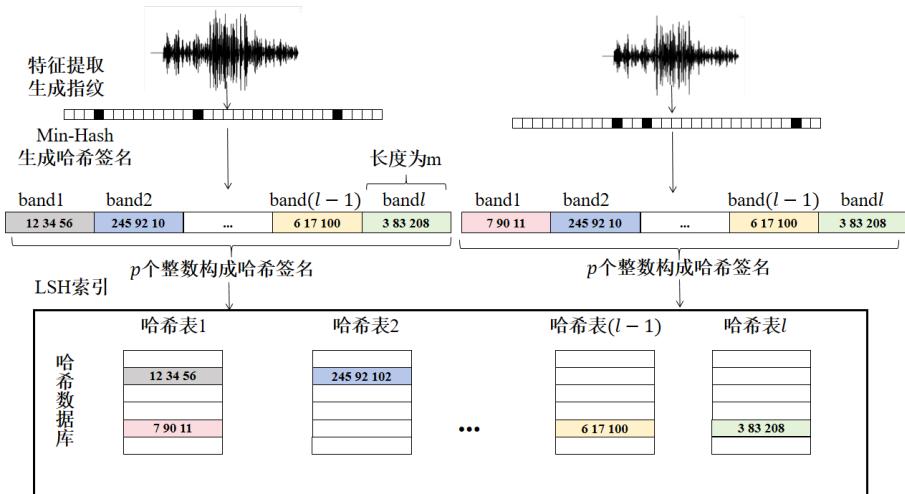


图 6 相似性搜索

Fig.6 Similarity Search Steps

相似性搜索步骤输出的是波形指纹对构成的相似度矩阵，如图 7 所示，由于连续地震波形会生成许多个指纹，因此相似指纹对数量也很庞大，需要通过后处理步骤进行指纹对的聚类。

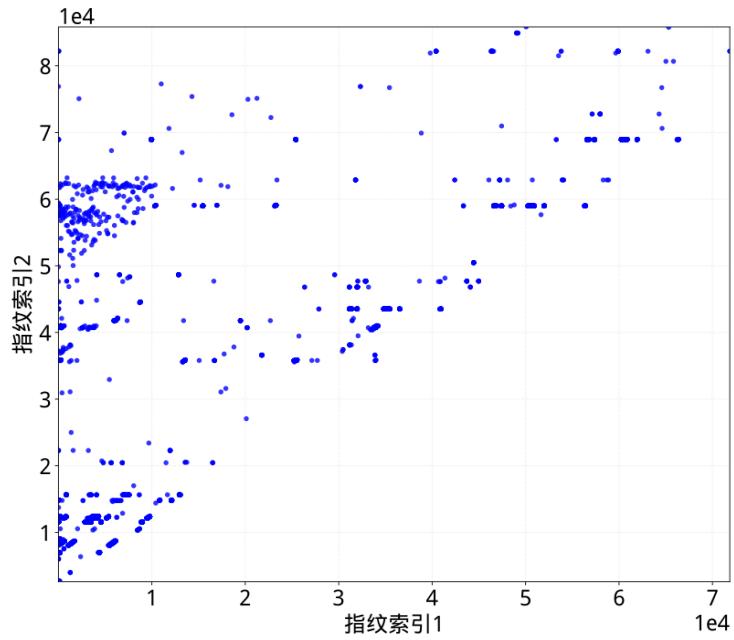


图 7 指纹对的相似度矩阵
Fig.7 Similarity matrix of fingerprint pairs

2.4 后处理

后处理模块通过三级结构化流程将海量指纹对提升至事件层级：

(1) **单台站事件提取**: 在相似度矩阵中识别沿对角线分布的高相似度簇——该模式表征具有连续时间关联性的相似指纹序列。基于事件持续时间的物理约束，设定时间邻近阈值合并重复检测窗口，生成初步台站级事件对列表。

(2) **多台站伪事件关联**: 利用重复地震事件的固定发震时刻差特性（同一台站的波形走时差异可忽略），设置台站检测阈值（记录到同一事件的最小台站数）。仅当多个台站独立检测到具有一致发震时刻差的事件对时，才关联构建候选事件。

(3) **事件解析与重建**: 融合同一物理事件的多重检测结果，基于综合置信度指标解析最终事件清单，显著压缩数据规模，提升了检测可靠性，克服了单指纹对仅依赖相似度指标的局限，且有效抑制了单台误检。

2.5 事件筛选

为提升检测结果的可信度，在最终处理阶段，基于核爆模板的先验信息及台站空间分布，引入时空约束与波形互相关算法对候选事件进行筛选：

(1) **时空约束**: 利用赤池信息准则 (Akaike Information Criterion, AIC) 方法优化检测信号初至(赵大鹏等, 2012)，该方法通过最小化时间序列分段模型的信息损失实现高精度到时检测：

$$AIC(k) = k \cdot \ln(\text{var}(x[1, k])) + (N - k - 1) \cdot \ln(\text{var}(x[k + 1, N])), \quad (5)$$

其中 k 表示假设的初至位置(第 k 个采样点)， x 为地震波形数据， N 为窗口长度， $\text{var}(x[1, k])$ 计算从第一个采样点到第 k 个采样点的方差，表示初至前背景噪声段的能量水平；

$\text{var}(x[k+1, N])$ 计算从第 $k+1$ 个采样点到第 N 个采样点的方差, 表示包含地震信号的能量水平, P 波初至时刻对应 $\text{AIC}(k)$ 的全局最小值点 k_{\min} 。通过对比震前和震后段信号的统计特征, AIC 方法能有效识别震相初至的突变点, 实现高精度自动到时拾取。依据区域速度模型计算监测区域内事件到周边各台站的理论到时差, 设定时空约束条件: 事件到各台站的实际观测到时差需在理论到时差允许的误差范围内, 剔除不满足该条件的事件。

(2) **波形互相关验证:** 以历史核爆事件在各台站的波形记录作为模板, 计算候选事件波形与对应台站模板波形的归一化互相关系数 (Cross-Correlation Coefficient, CC), 计算公式如下:

$$\text{CC} = \frac{\sum_{i=1}^n (T_i - \bar{T})(C_i - \bar{C})}{\sqrt{\sum_{i=1}^n (T_i - \bar{T})^2} \sqrt{\sum_{i=1}^n (C_i - \bar{C})^2}}, \quad (6)$$

其中, T_i 为模板波形的第 i 个采样点, C_i 为候选事件的第 i 个采样点, \bar{T} 与 \bar{C} 为对应波形的平均幅值, n 为波形采样点总数。保留互相关系数高于阈值的事件, 筛除与模板互相关系数偏离较大的误判事件。

3 结果分析

Seisprint 关键参数设置见表 2, 其中由于 MDJ 台站的采样率为 20Hz, 为了保持所有台站的参数一致, 其余 5 个台站统一重采样到 20Hz。

表 2 Seisprint 算法关键参数

Table 2 Key Parameters of the Seisprint Algorithm

Seisprint 参数名称	参数值
时频谱图窗口长度(s)	6
时频谱图窗口间隔(s)	0.2
采样率(Hz)	20
频谱图像长度(采样点数)	64
频谱图像间隔(采样点数)	5
频谱图像宽度(采样点数)	32
稀疏度参数 K	800
指纹长度	4096
哈希签名长度 p	400
哈希表个数 l	100
相似度阈值 ν	2
多台站关联阈值	4/6

接下来对影响算法性能的重点参数进行详细说明:

(1) 稀疏度参数 K 的选择: K 表示稀疏度参数, 即压缩后的非零小波系数的个数。K 决定了指纹的稀疏程度与判别细节量: 随着 K 增大, 指纹更稠密, 存储紧凑性下降, 相似度搜索步骤的运行时间增加; 随着 K 减小, 指纹更稀疏, 存储与检索更高效, 但可能牺牲对于弱信号或细粒度差异的敏感度。Bergen and Beroza (2019)指出, 对于 FAST 而言, 在 4096 维二进制指纹中, 取 $K=300-500$ (对应稀疏度 7-12%) 通常最优, 可在效率与灵敏度之间取得良好折中。本研究中的 Seisprint 与 FAST 的目标存在区别: FAST 更强调“检出微弱地震”;

Seisprint 除了要将核爆信号从噪声中稳健检出，还强调其与天然地震信号的鉴别性。为保留更多鉴别性细节，我们将稀疏度提高至约 20%，即 $K=800$ (4096 维)，该取值在指纹判别力与检索速度之间取得了较好的平衡（经数据集验证，见图 8）。

(2) 哈希表个数 l 与哈希签名长度 p 的选择：LSH 的精度和召回率主要通过这两个关键参数调节。由 2.3 小节可知，长度为 m 的 band 映射到 l 个独立哈希表；两个指纹在 m 个随机选取的非零位置上完全一致，才能被哈希到相同的哈希桶中。因此 m 越大，碰撞概率越低，候选项减少，召回率下降； m 越小，碰撞概率越大，召回率更高，但精度下降，且运行时间增加。为了在召回不下降的前提下提高精度，可增加哈希表个数 l ，但代价是更高的内存与索引维护成本。Rong et al. (2018)指出：哈希表个数 l 通常取 100， m 的经验建议为：短时长数据集（天-周）时 $m = 4$ ；较长的数据集（月-年） $m = 5$ ；最长数据集（5-10 年） $m = 6$ 。在本研究的数据集中，最长为 1 天的连续波形，因此取 $m = 4$ ($p = 400$)。

(3) 相似度阈值 v 的选择：为尽量保留潜在相似对，在相似性搜索阶段建议将初始阈值设得偏低，常用 $v = 2$ 作为起点，生成更充足的候选指纹集合；再于后处理中进一步修剪。

3.1 离线数据库测试

核爆事件自动检测的关键问题是需要具备鲁棒性和鉴别的双重优化性能。鲁棒性表示为在较低信噪比下将信号检出的能力，而鉴别性表示将核爆信号与非核爆信号区分的能力。为了验证 Seisprint 指纹提取算法的鉴别性，在离线事件库中，提取核爆、天然地震和背景噪声波形的指纹，选取 2016 年 1 月 6 日核爆事件 (NKT4) 波形作为模板，计算其余指纹与模板指纹相似度分布，如图 8 所示。其中，由于核爆事件和天然地震事件不一定可以同时关联到 6 个台站（例如 NKT1 仅关联到了 MDJ 台站），因此图 8 中不同台站的核爆事件数和天然地震事件数会出现差异。

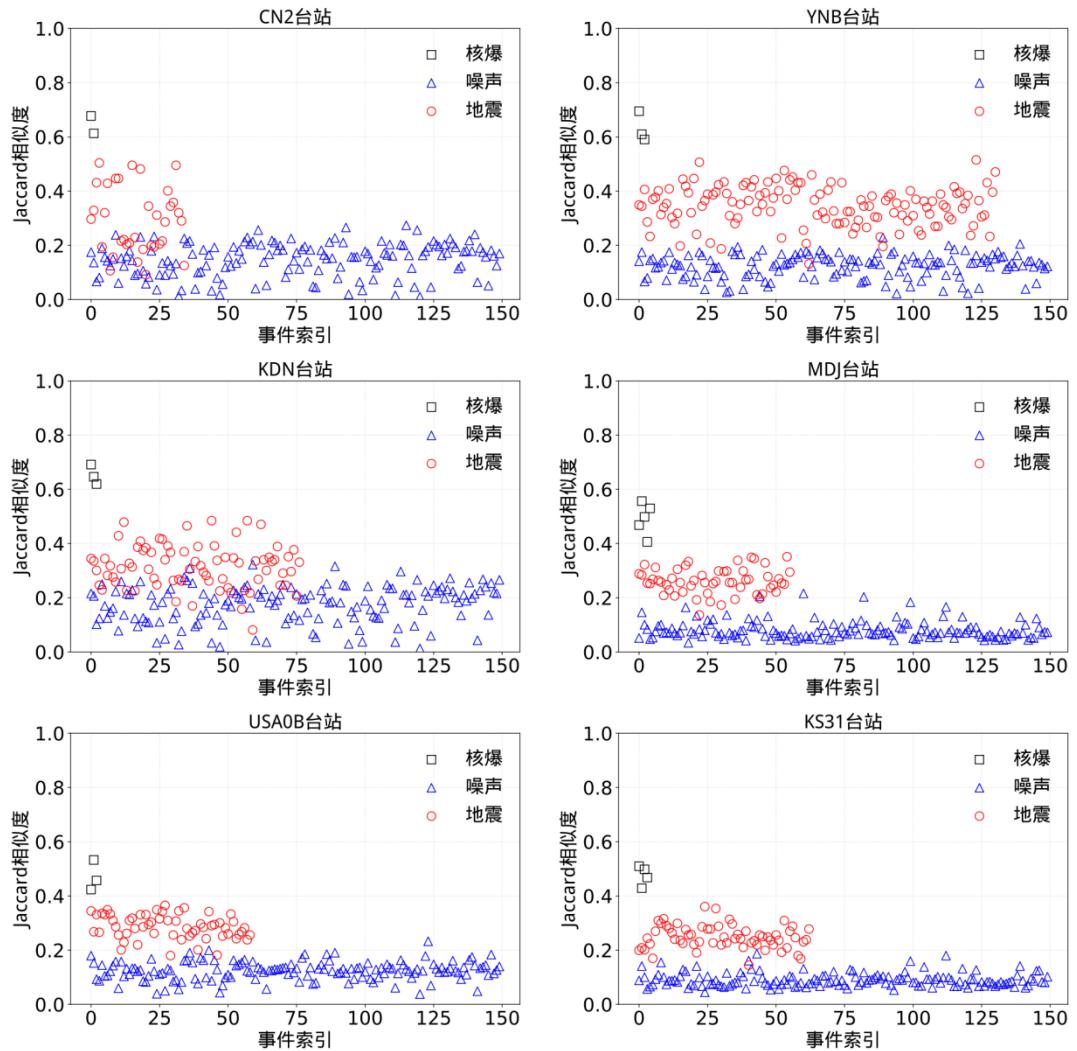


图 8 不同类型的指纹与模板指纹的相似度分布

Fig.8 Similarity Distribution Between Different Types of Fingerprints and the Template Fingerprint

显然，核爆指纹与模板指纹的相似度得分最高，天然地震指纹处于中间区间，与核爆指纹之间存在稳定的可分界；而背景噪声指纹得分普遍最低，基本集中在低相似度区域，形成明显的三层区分带；其中，天然地震与噪声的指纹区分度不高，甚至有的重叠在一起（如 CN2、KDN 台站），主要源于部分地震事件（见图 1）震级小于 2.5，信噪比低，地震信号被背景噪声淹没，导致其指纹与噪声指纹差异不显著。相反，震级较大且波形清晰的事件可保留更明显的特征，其指纹与噪声指纹差异明显。尽管各台站的匹配得分数值分布存在差异（可能源自于仪器响应、路径效应或背景噪声差异），但核爆、天然地震和背景噪声三类指纹在不同台站的分层趋势均一致。指纹空间中的聚类验证了以下特性：对于特定地区的事件，应用 Seisprint 指纹提取算法，相同震源类型（六次核爆炸）的波形生成高度相似的指纹特征，而不同震源类型（核爆炸与天然地震）则产生显著差异化的指纹，这说明 Seisprint 指纹提取算法对区分来自相同区域的不同震源类型的信号具有较好的鉴别性。

为评估 Seisprint 指纹提取算法对噪声干扰的鲁棒性，本研究采用向模板波形添加随机噪声的策略，并计算加噪指纹与模板指纹间的相似度。假设模板波形为 $\mathbf{x} \in R^M$ ，在时域中用 M 个样本表示，噪声波形 $\mathbf{n}^{(j)} \in R^M$ 取自同台站的噪声数据集， $F(\mathbf{x}) \rightarrow \{0,1\}^D$ 表示指纹提取操作，通过缩放因子 α 控制信噪比（SNR）：

$$\text{SNR} = 10 \log_{10} \left(\frac{\|\alpha \mathbf{x}\|^2}{\|\mathbf{n}\|^2} \right), \quad (7)$$

指纹精度Accuracy定义为加噪指纹与模板指纹的 Jaccard 相似度:

$$\text{Accuracy}(i) = J(F(\alpha_i \mathbf{x} + \mathbf{n}^{(i)}), F(\mathbf{x})), \quad (8)$$

基线相似度Baseline定义为噪声指纹与模板指纹间的 Jaccard 相似度:

$$\text{Baseline}(j) = J(F(\mathbf{n}^{(j)}), F(\mathbf{x})), \quad (9)$$

考虑到朝鲜地区历史核爆炸事件当量较大且选用近台数据,添加噪声对原始波形的影响可能不显著。为验证 Seisprint 算法对小当量核爆信号的检测鲁棒性,实验选用 KDN 台站记录的一次小震级天然地震事件波形为模板(事件参数与多台站波形见附件 2),对其添加不同信噪比的噪声进行测试。在 $\text{SNR} = 2 \text{ dB}$ 时(图 9),添加的噪声几乎淹没了有用信号,即使采用 1-2Hz 带通滤波,人工分析依旧存在困难。图 10 是应用 Seisprint 算法提取模板波形、加噪波形($\text{SNR} = 2 \text{ dB}$)与噪声的指纹,获得指纹精度和基线相似度的分布。在低信噪比条件下,指纹精度与基线相似度的分布依然呈现出较为明显的分界,位置大约在 Jaccard 相似度值为 0.2 附近。结果显示:对于低信噪比的地震波形记录,Seisprint 提取的指纹能够有效区分噪声和地震信号。

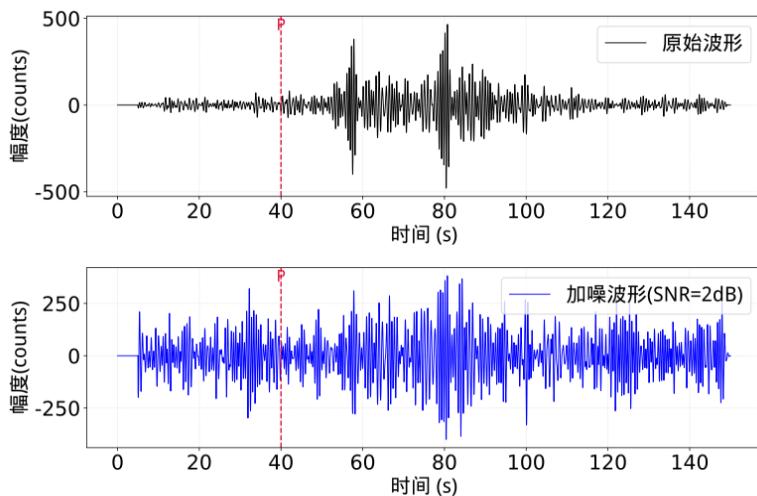


图 9 模板波形 (上) 与加噪后 $\text{SNR}=2\text{dB}$ 的波形 (下)

Fig.9 Original seismic waveform (top) and noise-added waveform with $\text{SNR}=2\text{dB}$ (bottom)

图中显示为 1-2Hz 滤波后的波形,红色虚线为 P 波到时

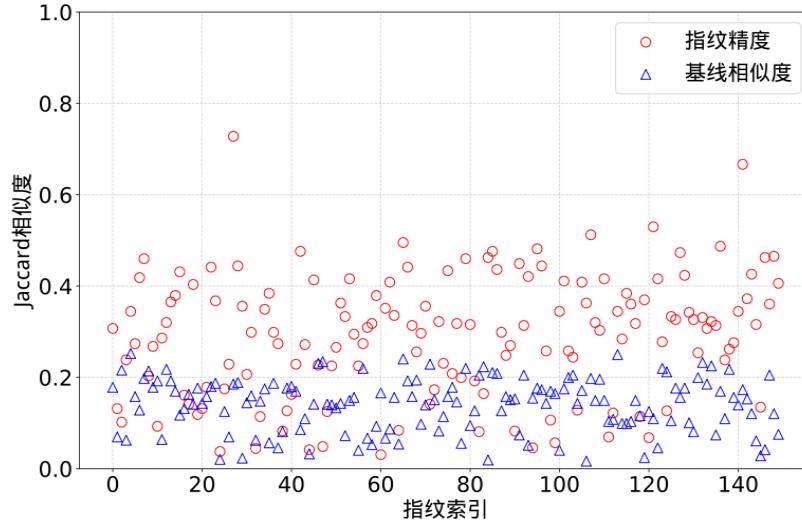


图 10 KDN 台站指纹精度与基线相似度分布 (SNR=2dB)

Fig.10 Distribution of fingerprint accuracy and baseline similarity at station KDN (SNR=2 dB)

在指纹提取的基础上，结合相似性搜索算法，我们进一步采用 Precision-Recall (PR) 曲线对 Seisprint 算法的检测性能进行综合评估。对于给定的相似度检测阈值，将模板指纹作为查询对象，利用 LSH 算法在包含加噪指纹和噪声指纹的数据库中进行快速相似性搜索。定义真阳性(True Positive, TP)为被正确识别为与模板指纹相似的加噪指纹，假阳性(False Positive, FP)为被错误识别为与模板指纹相似的噪声指纹。基于 TP 和 FP 的统计可给出精度(Precision)与召回率(Recall)这两个指标：

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (10)$$

$$\text{Recall} = \frac{TP}{N_{total}}, \quad (11)$$

其中， N_{total} 表示测试集中加噪指纹的总数量（即所有待检出的正样本数量）。精度衡量系统返回结果中正确匹配的比例，召回率衡量系统找出所有真正匹配项的能力。

PR 曲线描绘了当检测阈值在从最严格（相似度要求高）到最宽松（相似度要求低）的范围内连续变化时，系统所对应的 Precision 与 Recall 之间的关系。我们统计了不同 SNR 条件下的 PR 曲线，结果如图 11 所示。实验表明，当 $SNR \geq 5dB$ 时，在满足召回率大于 80% 的条件下，能够实现 90% 以上的检测精度。选择合适的检测阈值，系统能够较好地实现误检与漏检之间的权衡，这充分体现了 Seisprint 算法对噪声干扰的强鲁棒性。

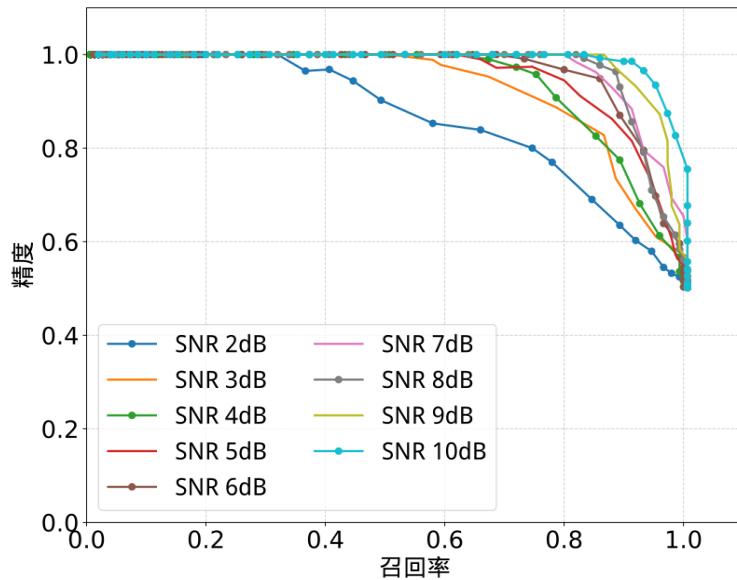


图 11 不同信噪比下的 PR 曲线

Fig.11 PR Curves under Different Signal-to-Noise Ratios

3.2 连续波形数据测试

为验证 Seisprint 系统在实时监测场景下的端到端检测效率与可靠性，开展基于连续地震波形数据的测试。选取离线数据库中 2016 年 1 月 6 日朝鲜核爆事件的波形片段，构建“帧头”，将其拼接至对应台站的连续波形数据起始端，以驱动后续的核爆指纹匹配检测流程。如图 3 的检测流程，根据是否与帧头匹配，可以将事件分为波形相似的天然地震事件和候选事件，对于候选事件，算法启动事件筛选程序，满足筛选条件的则触发警报，表明监测区域内可能存在核爆炸事件，需由分析员进一步人工复核，否则不触发警报。本文利用 Seisprint 算法对 24 天的连续波形进行测试，共检测出 22 个事件，包括 16 次波形相似的天然地震事件和 6 次候选事件，经过事件筛选，候选事件中触发 3 次警报。

针对这 3 次报警事件和另外 3 次未触发警报的候选事件，绘制其在多台站的波形对比图（图 12）。采用 AIC 算法检测信号初至，报警事件在 MDJ、CN2 等 5 个台站的初至信号到时差不超过 10 s，并且 YNB 台检测到的朝鲜核试验场的初至信号到时都早于其他台站（图 2 中显示 YNB 台站距离监测区域最近，因此理论走时最短）。相比之下，其他 3 个候选事件在这些地震台站记录上的初至信号到时差不具备上述特征，通过人工分析并与标准地震目录对比，确定候选事件均为监测区域外的天然地震事件，详细事件参数见附件 3。

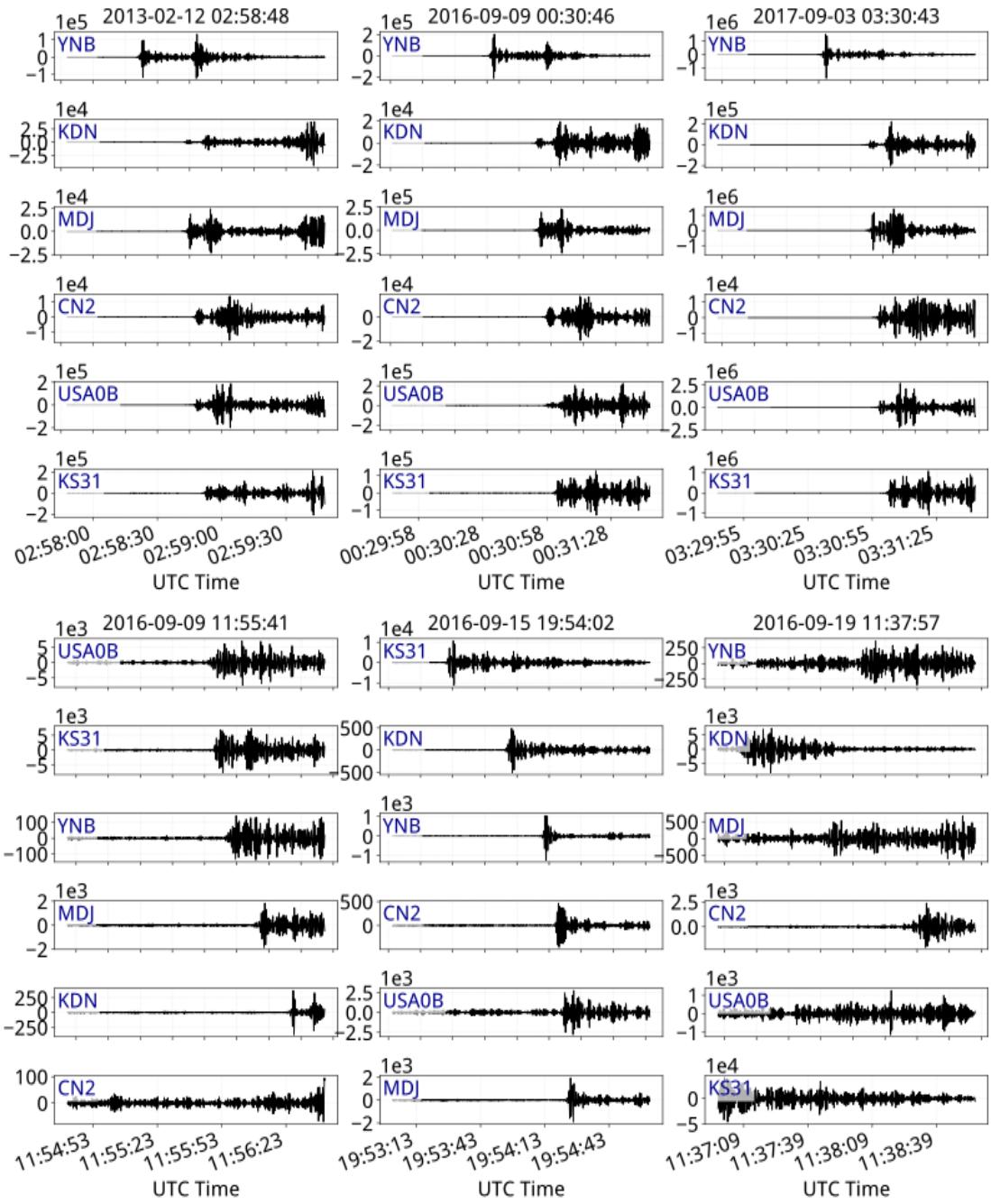


图 12 检测事件的多台站波形

Fig.12 Multi-station Waveforms of Detected Events

(a)-(c)为报警事件, (d)-(f)为其余候选事件

为进一步验证检测结果, 对上述 6 个候选事件在多个台站分别截取检测时刻 ± 60 s 的波形段, 与帧头波形进行互相关计算, 并提取最大互相关系数 CC (表 3)。设定判别标准为: 在 6 个参与台站中, 需满足 ≥ 5 个台站的 $CC \geq 0.5$, 才会判定为核爆事件。结果显示, 真实核爆事件全部满足判定标准, 而候选事件的平均 CC 仅 0.20–0.27, 且所有台站数据的互相关系数 $CC < 0.5$ 。通过多台站的初至信号到时差与波形互相关分析, 可确认 3 个候选事件不符合报警条件。

表 3 检测事件多台站互相关检测结果

Table 3 Cross-Correlation Detection Results of Events at Multiple Stations

事件时间	MDJ	CN2	KS31	USA0B	YNB	KDN	平均互相关系数	判定结果
2013/02/12:02:58:56	0.72	0.66	0.49	0.63	0.68	0.62	0.63	真实核爆
2016/09/09/00:30:46	0.91	0.76	0.67	0.25	0.90	0.67	0.69	真实核爆
2016/09/09/11:56:38	0.20	0.20	0.22	0.20	0.18	0.32	0.22	误判
2016/09/15/19:54:32	0.22	0.32	0.25	0.23	0.28	0.33	0.27	误判
2016/09/19/11:37:57	0.15	0.21	0.26	0.12	0.17	0.26	0.195	误判
2017/09/03/03:30:43	0.63	0.58	0.62	0.57	0.62	0.61	0.605	真实核爆

表 4 给出了 Seisprint、NET-VISA(王娟等, 2021)和波形互相关方法对于 NKT3、NKT5 和 NKT6 当天连续波形数据的检测结果，并与 CTBTO REB 标准目录进行了对比。其中 Seisprint 和波形互相关均采用 2016 年 1 月 6 日朝鲜核爆炸事件 (NKT4) 的波形作为模板，波形互相关方法参数设置如下：带通滤波频段为 0.8–4 Hz，模板长度 20 s，单台站检测 (MDJ 台站)，采样率 20 Hz，相关阈值 0.5。检测结果显示，三种方法均成功检出三次核爆事件，NET-VISA 将 NKT5 和 NKT6 均分裂成了两个独立事件，波形互相关的结果包含 2 次误检。

在时间误差方面，Seisprint 相较于 NET-VISA 和波形互相关更大。这是因为 Seisprint 的检测流程中未包含震相拾取步骤。在后处理阶段的单台站事件提取步骤中，Seisprint 设定将指纹合并的最大时间间隔为 P 波与 S 波的最大到时差（通常为 15–40 s），并以相似度得分最高的指纹对应时间作为事件检测时间。在本研究中，指纹长度为 10 s，因此理论上 Seisprint 的检测时间相较真实发震时刻的误差最大可达约 50 s。

如果期望得到精确的发震时刻，在 Seisprint 流程之后可考虑添加震相拾取和定位步骤。例如，可截取检测时间前后各 60 s 的波形，利用匹配-定位 (M&L) 方法或机器学习方法进行震相识别与定位，给出事件的精确时刻与位置。

表 4 三种方法的事件检测结果

Table 4 Event detection results of the three methods

检测方法	日期	时间	时间误差
Seisprint	2013-02-12	02:57:56	+5s
	2016-09-09	00:30:46	+45s
	2017-09-03	03:30:43	+42s
NET-VISA	2013-02-12	02:57:50	-1s
	2016-09-09	00:30:04	+3s
	2016-09-09	00:30:12	/
	2017-09-03	03:30:01	0
波形互相关	2013-02-12	03:30:05	/
	2013-02-12	23:58:34	/
	2013-02-12	02:58:00	+9s
	2016-09-09	00:30:11	+10s
	2017-09-03	03:30:09	+8s
	2017-09-03	23:58:15	/

对于 NKT6 后发生的三次余震事件 PEV1-PEV3，Seisprint 算法均未检出。已有研究(He et al., 2018)指出，这些事后事件在主频带与各相位 (P、S、Rg/Lg) 相对幅度关系上显著区

别于核爆，提示其震源时间函数与机制不同：其中 PEV1 很可能为爆炸腔体坍塌或诱发构造地震，而 PEV2、PEV3 更可能属于构造型事件(Yao et al., 2018; 刘俊清等, 2019)。采用 Seisprint 方法对 PEV1–PEV3 的指纹提取过程与详细分析见附件 4。因此，连续波形数据测试结果表明：Seisprint 对核爆事件和天然地震事件的鉴别性较好，它能稳定检出核爆事件的同时，不会将同一区域内的天然地震事件错误检出。

3.3 计算效率

为了评估 Seisprint 方法在实际应用中的计算效率，我们对其内存需求与运行时间进行了分析。其内存占用主要由以下三部分构成：

(1) 哈希表指针开销 $O(l * M)$ ：其中 M 为指纹数量， l 为哈希表数量。每个哈希表中存储指向子指纹的指针，每个指针占用 4 字节。

(2) 哈希桶结构开销 $O(l * b)$ ：其中， b 为每个哈希表中哈希桶的数量。我们采用数组实现哈希表，每个哈希桶存储一个指向其内容链表的指针（4 字节）和一个记录桶内子指纹数量的整数（4 字节），共计 8 字节。

(3) 子指纹存储开销 $O(p * M)$ ：其中 p 为哈希签名长度。

在本文的测试场景中，参数设定如下：每秒生成一个指纹，处理 24 小时波形数据并拼接 150 秒帧头，总指纹数量 $M = 86550$ 。哈希表数量 $l = 100$ ，子指纹长度（哈希签名长度） $p = 400$ ，每个哈希表的哈希桶数量 $b = 100,000$ 。据此计算各部分内存开销：

$$O(l * M): 4 \text{ 字节} * 100 * 86550 = 34,620,000 \text{ 字节} \approx 33 \text{ MB}, \quad (12)$$

$$O(l * b): 8 \text{ 字节} * 100 * 100,000 = 80,000,000 \text{ 字节} \approx 80 \text{ MB}, \quad (13)$$

$$O(p * M): 1 \text{ 字节} * 400 * 86550 = 34,620,000 \text{ 字节} \approx 33 \text{ MB}, \quad (14)$$

假设额外开销约为 1MB。那么内存需求总计约为 147MB。扩展至 1 年（365 天）的内存需求约为 53GB，此内存规模在普通服务器或集群环境中可行。Seisprint 的计算复杂度与 FAST 相似，特征提取与数据持续时间呈线性关系，相似性搜索为近线性复杂度 $O(M^{1.36})$ 。经统计，在 Linux 环境（Intel(R) Xeon(R) Gold 6248R CPU @ 3.00GHz, x86-64, 96CPUs, 1.5MB 缓存）下，处理 6 个台站的 24 小时连续波形数据的端到端耗时均在 1 分 30 秒以内。

当模板数增加时，Seisprint 算法只需要将新的模板继续拼接至帧头，运行时间只与新增模板自身长度相关；相对于大规模连续波形，模板极短，因此带来的开销可忽略不计；而每增加一个模板，波形互相关算法的运行时间都需要增加一倍，当扩展到大规模连续数据时，Seisprint 算法的计算效率要远远高于波形互相关。

4 结论

为监测特定地区的核试验活动，本研究提出了一种基于波形指纹的核爆炸事件快速检测方法 Seisprint。该方法的核心流程包括：选取历史核爆炸波形作为模板帧头，拼接至连续波形数据起始处；将波形数据压缩为稀疏二进制指纹，借助哈希数据库实现快速相似性搜索；再结合时空约束与互相关阈值对候选事件进行严格筛选，仅当多个台站均检测到与模板波形相似且符合判别条件的事件时才触发警报。利用朝鲜核试验场周边 6 个台站记录的 6 次历史

核爆及天然地震数据进行验证，结果表明：Seisprint 可在 2 分钟内完成多台站 24 小时连续数据的处理，实现了对该地区核爆炸事件的快速准确识别。与国际数据中心现有波形处理算法相比，本方法在特定地区的应用中表现出显著优势：运行效率高、内存占用低、具有良好的可扩展性，同时保持较高的检测准确性。通过替换模板帧头，该方法还可推广至其他地区的核爆炸或重复地震事件的检测中。

然而，Seisprint 在指纹提取过程中对信号时频谱进行了多尺度边缘与纹理压缩，将丰富的能量分布信息转换为二值化轮廓，导致一定程度的信息损失。区域内不同类型地震事件因震源机制差异而呈现不同波形特征，当前指纹表征方式在区分细微特征方面尚存在局限。下一步改进将聚焦于更精细的波形指纹特征提取方法，并结合更高效的模式匹配算法，以提升对相似事件的区分能力与检测可靠性。

作者贡献：

弓妮（第一作者）：负责完成了创新点的提出、研究方案设计、算法实现和应用测试、数据分析与论文撰写工作。

商杰：在数据预处理、算法优化及相关测试方面做出关键贡献。

唐伟：负责论文背景指导，并参与测试结果分析。

刘哲函：参与系统性能评估工作。

王海军（通讯作者）：算法原理分析和应用指导、论文修改和审阅。

黄立洪：协助构建测试数据集。

韩守诚：参与背景调研与参考文献整理。

江宇：对文章结构、语言表达进行润色与审校。

Reference

- Ali, S. M., Qorbani, E., Le Bras, R., et al., 2025. Interactive Analysis of the Results of Net-Visa, a Bayesian Inference System, in Ctbto's International Data Centre Bulletin Production. *Acta Geophysica*, 73(1), 71-81. <https://doi.org/10.1007/s11600-024-01398-0>
- Arora, N. S. 2012. Model-Based Bayesian Seismic Monitoring - Escholarship.
- Arora, N. S., Russell, S. J., Sudderth, E. B. 2009. Vertically Integrated Seismological Analysis Ii : Inference.
- Arora, N. S., Russell, S. J., Sudderth, E. B., 2013. Net - Visa: Network Processing Vertically Integrated Seismic Analysis. *Bulletin of the Seismological Society of America*, 103(709-729).
- Asada, M., 2002. Ctbt: Legal Questions Arising from Its Non - Entry - into - Force. *Journal of Conflict and Security Law*, 7(1), 85-122. <https://doi.org/10.1093/jcls/7.1.85>
- Bauer, S., O'Reilly, C. 2016. The Comprehensive Nuclear-Test-Ban Treaty Organization (Ctbt): Current and Future Role in the Verification Regime of the Nuclear-Test-Ban Treaty. Nuclear Non-Proliferation in International Law: Volume II - Verification and Compliance. J. L. Black-Branch and D. Fleck. The Hague, T.M.C. Asser Press: 131-150.
- Bergen, K. J., Beroza, G. C., 2019. Earthquake Fingerprints: Extracting Waveform Features for Similarity-Based Earthquake Detection. *Pure and Applied Geophysics*, 176(3), 1037-1059. <https://doi.org/10.1007/s00024-018-1995-6>
- Bobrov, D., Kitov, I., 2011. Analysis of the 2008 Chinese Earthquake Aftershocks Using Cross-Correlation. *Proceedings of the 2011 Monitoring Research Review: 'Ground-Based Nuclear Monitoring Technologies'*, 811-821.
- Bobrov, D., Kitov, I., Zerbo, L., 2014. Perspectives of Cross-Correlation in Seismic Monitoring at the International Data Centre. *Pure and Applied Geophysics*, 171(3), 439-468. <https://doi.org/10.1007/s00024-012-0626-x>
- Bobrov, D. I., Kitov, I. O., Rozhkov, M. V., et al., 2016. Towards Global Seismic Monitoring of Underground Nuclear Explosions Using Waveform Cross Correlation. Part II: Synthetic Master Events. *Seismic Instruments*, 52(3), 207-223. <https://doi.org/10.3103/S0747923916030038>
- Bras, R. L., Russell, S. J., Arora, N. S., et al. 2011. Machine Learning at the Ctbto . Testing , and Evaluation of the False Events Identification (Fei) and Vertically Integrated Seismic Association (Visa) Projects.
- Cohen, E., Datar, M., Fujiwara, S., et al. 2000. Finding Interesting Associations without Support Pruning. *Proceedings of 16th International Conference on Data Engineering (Cat. No.00CB37073)*.
- Coyne, J., Bobrov, D., Bormann, P., et al. 2012. Ctbt: Goals, Networks, Data Analysis and Data Availability. Postdam, Germany :, GFZ.
- Gibbons, S. J., Ringdal, F., 2006. The Detection of Low Magnitude Seismic Events Using Array-Based Waveform Correlation. *Geophysical Journal International*, 165(1), 149-166. <https://doi.org/https://doi.org/10.1111/j.1365-246X.2006.02865.x>
- Gibbons, S. J., Ringdal, F., 2012. Seismic Monitoring of the North Korea Nuclear Test Site Using a Multichannel Correlation Detector. *IEEE Transactions on Geoscience and Remote Sensing*, 50(5), 1897-1909. <https://doi.org/10.1109/TGRS.2011.2170429>
- Gionis, A., Indyk, P., Motwani, R. 1999. Similarity Search in High Dimensions Via Hashing. *Proceedings of the 25th International Conference on Very Large Data Bases*, Morgan Kaufmann

- Publishers Inc.: 518–529.
- He, X., Zhao, L. F., Xie, X. B., et al., 2018. High - Precision Relocation and Event Discrimination for the 3 September 2017 Underground Nuclear Explosion and Subsequent Seismic Events at the North Korean Test Site. *Seismological Research Letters*, 89(6), 2042-2048. <https://doi.org/10.1785/0220180164>
- Kushida, N., Mialle, P., Tomuta, E., et al. 2017. The Netvisa Automatic Association Tool. Next Generation Software Testing and Performance under Realistic Conditions.
- Le Bras, R., Arora, N., Kushida, N., et al., 2021. Net-Visa from Cradle to Adulthood. A Machine-Learning Tool for Seismo-Acoustic Automatic Association. *Pure and Applied Geophysics*, 178(7), 2437-2458. <https://doi.org/10.1007/s00024-020-02508-x>
- Le Bras, R., Swanger, H., Sereno, T., et al., 1994. Global Association. 38.
- Li, Z.,Zhan, Z., 2018. Pushing the Limit of Earthquake Detection with Distributed Acoustic Sensing and Template Matching: A Case Study at the Brady Geothermal Field. *Geophysical Journal International*, 215(1583-1593). <https://doi.org/10.1093/gji/ggy359>
- LIU, J.Q., LIU, C., ZHANG, Y., Yu, et al., 2019. The moment tensor of ML3.4 earthquake on September 23, 2017, Democratic People's Republic of Korea (DPRK). *Chinese Journal of Geophysics (in Chinese)*, 62(2): 534-543.
- Muir, J., Fernando, B.,Barrett, E., 2023. False Positives Are Common in Single-Station Template Matching. *Seismica*, 2(2), <https://doi.org/10.26443/seismica.v2i2.385>
- Rong, K., Yoon, C. E., Bergen, K. J., et al., 2018. Locality-Sensitive Hashing for Earthquake Detection: A Case Study of Scaling Data-Driven Science. *Proc. VLDB Endow.*, 11(11), 1674–1687. <https://doi.org/10.14778/3236187.3236214>
- Senobari, N. S., Funnyng, G. J., Keogh, E., et al., 2018. Super - Efficient Cross - Correlation (Sec - C): A Fast Matched Filtering Code Suitable for Desktop Computers. *Seismological Research Letters*, 90(1), 322-334. <https://doi.org/10.1785/0220180122>
- Shaheen, S., Khanum, A., Khan, S. A., et al., 2015. Using Mac Frame Header for Efficient Multimedia Streaming over Ieee 802.11 Wireless Lan. *2015 International Conference on Computer, Communications, and Control Technology (I4CT)*, 570-574.
- Stéphane, M. 2009. Chapter 7 - Wavelet Bases. *A Wavelet Tour of Signal Processing (Third Edition)*. M. Stéphane. Boston, Academic Press: 263-376.
- Sugita, D., Ito, Y. 2015. Study of Web Service Estimation Using Only Ethernet Frame Header or Ip Packet Header. *2015 IEEE 4th Global Conference on Consumer Electronics (GCCE)*.
- Wang, J., Liu, J.M., Li, J., et al., 2021. Progress and testing of NET-VISA software[J]. *Progress in Earthquake Sciences(in Chinese)*, 51(6): 274-278.
- Yao, J., Tian, D., Lu, Z., et al., 2018. Triggered Seismicity after North Korea's 3 September 2017 Nuclear Test. *Seismological Research Letters*, 89(6), 2085-2093. <https://doi.org/10.1785/0220180135>
- Yoon, C. E., O'Reilly, O., Bergen, K. J., et al., 2015. Earthquake Detection through Computationally Efficient Similarity Search. *Sci Adv*, 1(11), e1501057. <https://doi.org/10.1126/sciadv.1501057>
- Zhao, D.P, Liu, X.Q., Li, H., et al., 2012. Application of kurtosis and AIC methods in automatic identification of regional seismic events and direct P-wave first motions. *Journal of Seismological Research(in Chinese)*, 35(2): 220-225.
- Zhao, L.F., Xie, X.B., He, X., et al. 2017. Seismological identification and yield estimation of the September 3, 2017 North Korean underground nuclear test. *Chinese Science Bulletin(in Chinese)*,

2017, 62(35): 4163-4168.

中文参考文献

- 刘俊清, 刘财, 张宇, 等, 2019. 2017 年 9 月 23 日朝鲜 ML3.4 地震矩张量反演. 地球物理学报, 62(2): 534-543.
- 王娟, 刘俊民, 李健, 等, 2021. NET-VISA 软件研发进展与性能测试, 地震科学进展, 51 (6): 274-278.
- 赵大鹏, 刘希强, 李红, 等, 2012. 峰度和 AIC 方法在区域地震事件和直达 p 波初动自动识别方面的应用. 地震研究, 35(2), 220-225.
- 赵连锋, 谢小碧, 何熹, 等, 2017. 2017 年 9 月 3 日朝鲜地下核试验的地震学鉴别和当量估计. 科学通报, 2017, 62(35):4163-4168.

附件 1

Haar 小波变换是最简单的一种正交小波变换，具有快速、结构简单的特点，常用于图像压缩、边缘提取、特征降维等任务。本文在时频谱图（2D 图像）上提取小波系数，下面给出 Haar 小波变换的一维和二维离散形式公式：

1.一维 Haar 小波变换公式（离散）

设输入信号为长度为 $N = 2^k$ 的一维序列：

$$\mathbf{x} = [x_0, x_1, x_2, \dots, x_{N-1}], \quad (1)$$

定义 Haar 小波的低频滤波器（尺度函数）：

$$h = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right], \quad (2)$$

高频滤波器（小波函数）：

$$g = \left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right], \quad (3)$$

计算方式如下（对每两个相邻采样值进行变换），近似系数（低频分量）：

$$a_k = \frac{x_{2k} + x_{2k+1}}{\sqrt{2}}, \quad (4)$$

细节系数（高频分量）：

$$d_k = \frac{x_{2k} - x_{2k+1}}{\sqrt{2}}, \quad (5)$$

其中， $k = 0, 1, \dots, \frac{N}{2} - 1$ 。

2.二维 Haar 小波变换公式

对二维图像（例如时频谱图） $I[m, n]$ 应用 Haar 小波变换时，过程如下：

(1) 先对每一行做一维 Haar 变换，得到：低频行系数 $L[m, n]$ 和高频行系数 $H[m, n]$ ；

(2) 再对每一列分别做一维 Haar 变换，得到四个子图块：

LL：低频-低频（近似图像）

$$LL = \frac{I_{2m,2n} + I_{2m+1,2n} + I_{2m,2n+1} + I_{2m+1,2n+1}}{2}, \quad (6)$$

LH：低频-高频（水平边缘）

$$LH = \frac{I_{2m,2n} - I_{2m+1,2n} + I_{2m,2n+1} - I_{2m+1,2n+1}}{2}, \quad (7)$$

HL：高频-低频（垂直边缘）

$$HL = \frac{I_{2m,2n} + I_{2m+1,2n} - I_{2m,2n+1} - I_{2m+1,2n+1}}{2}, \quad (8)$$

HH：高频-高频（对角边缘）

$$HH = \frac{I_{2m,2n} - I_{2m+1,2n} - I_{2m,2n+1} + I_{2m+1,2n+1}}{2}, \quad (9)$$

其中， m, n 为输出图像的行列索引。

因此，Haar 小波通过对局部区域的加减操作，提取出信号（或图像）中的近似和平滑变化（低频）与细节变化（高频）；对 STFT 生成的时频谱图进行 Haar 2D 变换可提取出不同方向（水平、垂直、对角）的频率变化特征，有助于后续分类或检索。

附件 2

为验证 Seisprint 算法对小当量核爆信号的检测鲁棒性，实验选用 KDN 台站记录的一次小震级天然地震事件波形为模板，其中事件参数见表 S1，多台站波形见图 S1。

表 S1 模板事件参数

Table S1 Parameters of Template Event

	日期	UTC 时间	纬度	经度	震级(mb)
模板事件	2018/02/07	21:46:03	41.37	129.01	3.7

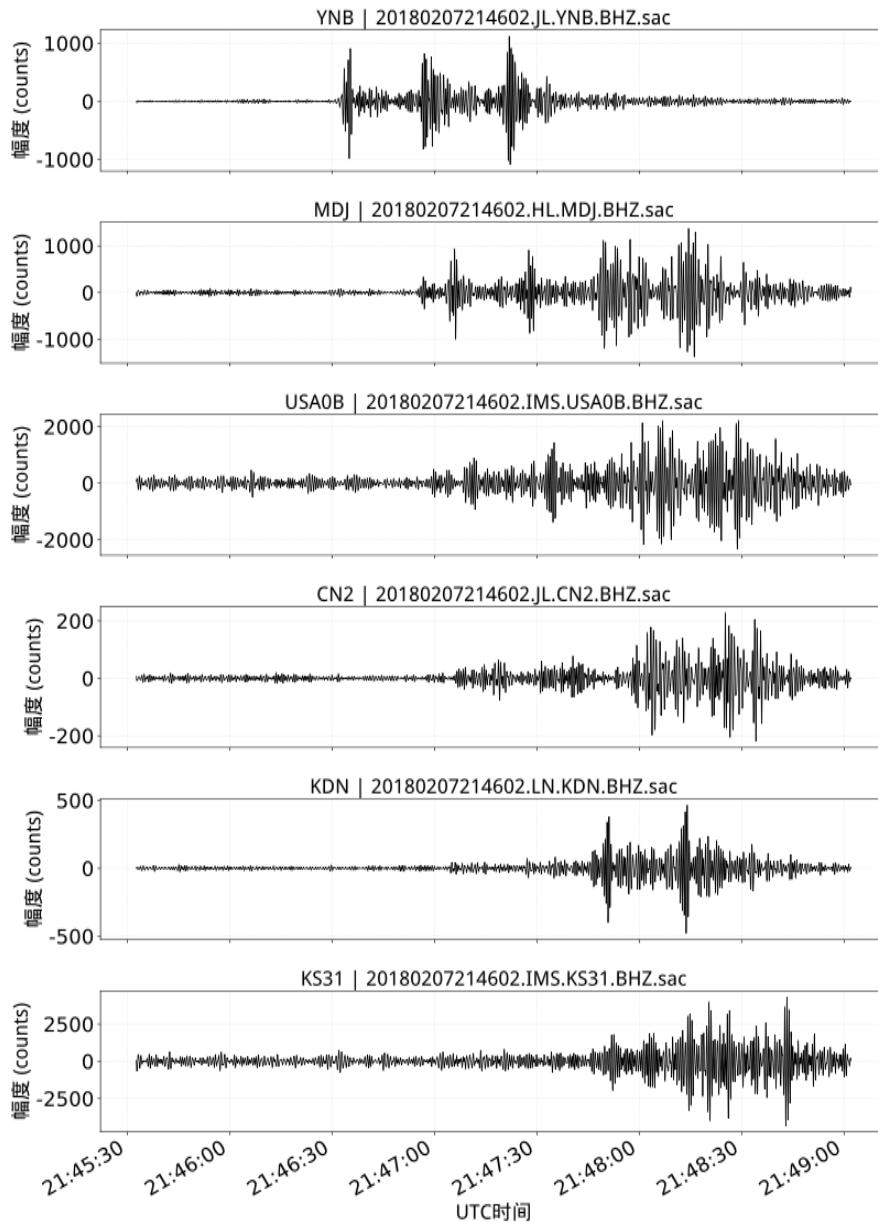


图 S1 模板事件在多个台站的波形

Fig. S1 Waveforms of Template Event at Multiple Stations

附件 3

我们经过人工分析并与标准地震目录进行对比，可以确认连续波形数据测试得到的三个未报警的候选事件均为监测区域外的天然地震事件，事件参数见表 S2。

表 S2 候选事件参数

Table S2 Parameters of Candidate Events

候选事件	日期	UTC 时间	纬度	经度	震级(mb)
事件 1	2016/09/09	11:53:17	36.30	140.97	4.82
事件 2	2016/09/15	19:45:57	-3.85	129.04	5.61
事件 3	2016/09/19	11:34:03	35.86	128.99	3.76

附件 4

图 S2 给出了对三次余震事件(PEV1–PEV3)采用 Seisprint 方法进行指纹提取的流程，将其与图 5 进行对照，可以发现这三次的余震的时频谱图(图 S2(d)–(f))在频带能量分布与相位特征上表现出与天然地震更接近；尽管从原始波形上看，PEV1 与 PEV3 更相似，但是将时域波形转化为时频谱图后，比起 PEV1，PEV2 与 PEV3 的时频谱更相似。这与 He et al. (2018) 的分析一致：PEV1 更可能为爆炸腔体坍塌或诱发事件，而 PEV2、PEV3 更接近构造型地震。

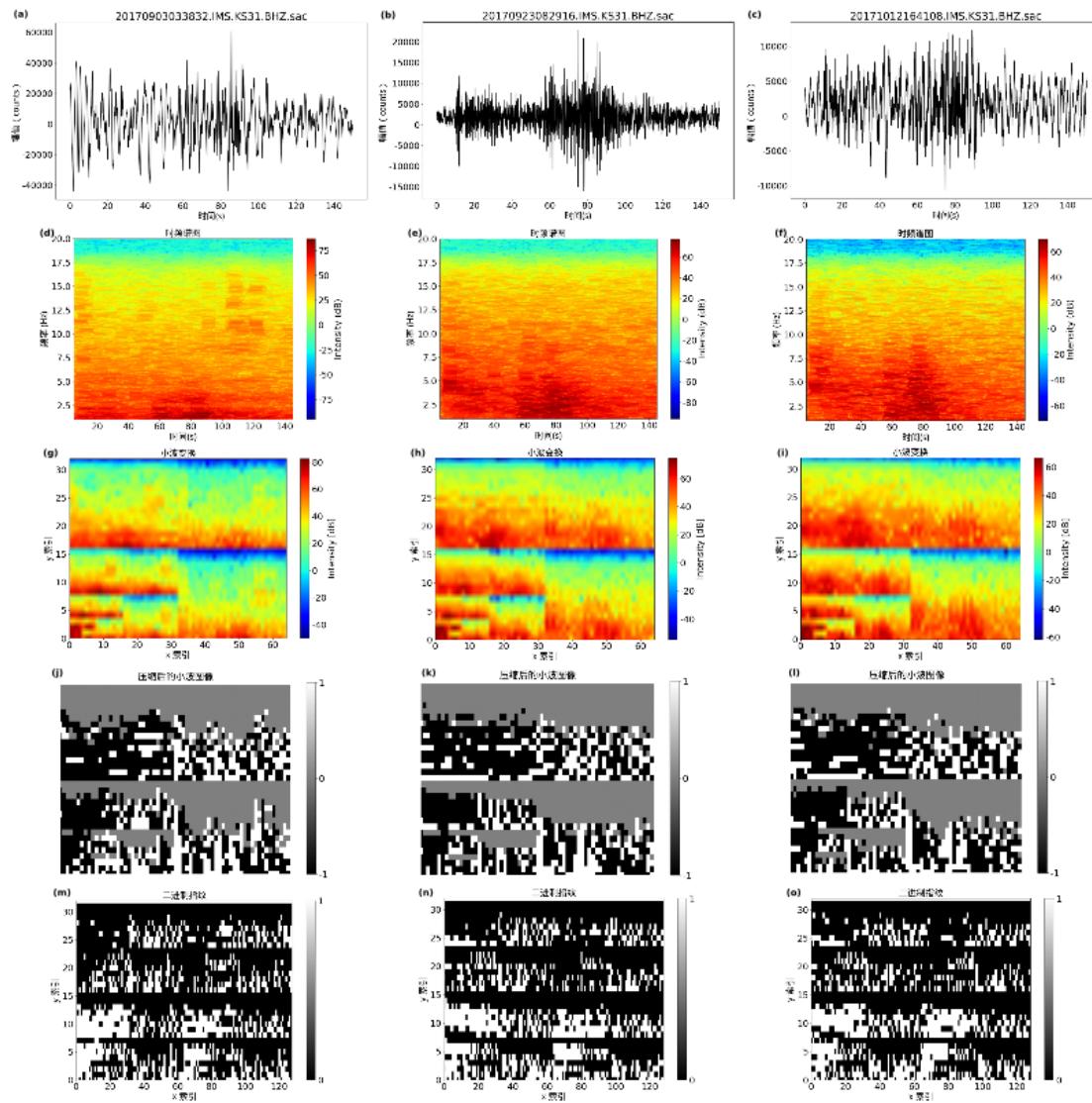


图 S2 Seisprint 指纹提取算法生成三次余震事件 (PEV1–PEV3) 的指纹

Fig S2. Fingerprint extraction steps for three aftershock events (PEV1–PEV3)

基于 Seisprint 的指纹提取结果，我们分别计算了核爆模板 (NKT4) 与其他核爆事件、及 PEV1–PEV3 的指纹之间的 Jaccard 相似度(见图 S3)。可以看到，“核爆–核爆”的指纹相似度整体显著高于“核爆–余震”，两类分布之间存在清晰分界。因而在与正文一致的检索阈值设置下，三次余震的指纹相似度未达到触发阈值，未被 Seisprint 检出为候选事件。该结果与震源类型判别相互印证，进一步说明 Seisprint 在保障核爆稳定检出的同时，能够有效避免将同一区域内的天然地震误报为核爆。

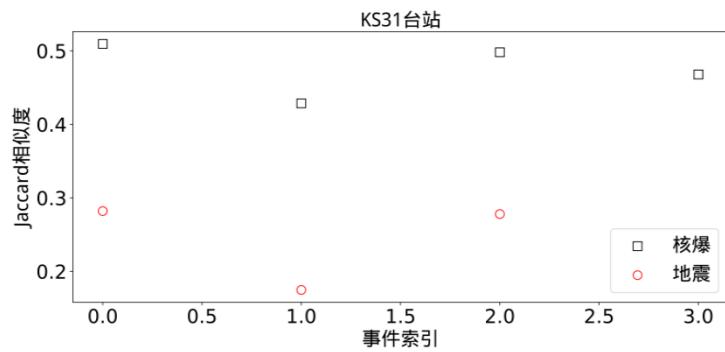


图 S3 核爆指纹、余震指纹与模板指纹的相似度分布

Fig S3. Similarity distributions among nuclear-explosion, aftershock, and template fingerprints.