

地震科技前沿快报

2024 年第 4 期（月刊总第 63 期）

中国地震局科学技术委员会
中国地震局科技与国际合作司

中国科学院兰州文献情报中心
中国科学技术信息研究所

本期概要

1、结合微震记录和大地测量评价华北地区的地震危险性

华北地区位于亚洲大陆东部，在历史上和近期经历了多次极具破坏性的地震，造成了巨大的生命财产损失。来自中国的研究人员通过将地震记录和大地测量相结合来综合评价华北地区的地震潜力。结果表明，较大应变速率主要集中在鄂尔多斯地块西缘和唐山断裂带。不同的应变速率参数所产生地震重现期（recurrence interval）的空间分布比较相似，但大小不同。与华北其他地区相比，鄂尔多斯地块西缘和唐山断裂带呈现出较短（7000~15000 年）的 M_L 7.5 级地震重现期。

2、利用大数据和人工智能进行地震预报：一项为期 30 周的中国实时案例研究

前兆信号探测是地震预报的关键之一，人工智能（AI）技术以全新的方式打开了检测地震前兆信号的大门。研究人员使用多分量地震监测系统 AETA 提供的电磁（electromagnetic, EM）和地声（geo-acoustic, GA）数据来预测中国的地震，并在中国西南地震区进行了测试。结果表明，测试准确率达到 70%，而精确度、召回率分别为 63.63%、93.33%。与目录解相比，使用该方法预测的距离和震级的平均绝对误差分别为 381 km 和 0.49 级。

3、利用梯度提升回归器根据 P 波特征实时预测距离和 PGA 在现

地预警中的应用

现地预警（on-site EEW）系统是减少地震灾害的重要手段。来自意大利的研究人员探索了使用机器学习回归实时预测震中距离（D）和峰值地面加速度（PGA）的可行性，并提出了一个基于 D 和 PGA 预测的概率现地预警系统模型，测试分析结果表明，所提出的系统在 1 s 窗口的情况下是可靠的，这对于意大利中部这样的地震易发地区（具有中到大地震风险）来说具有巨大优势。

结合微震记录和大地测量评价华北地区的地震危险性

中国华北地区在历史上和近期遭受了多次极具破坏性的地震，造成了巨大的生命财产损失。因此，确定未来最有可能发生大地震的地区是至关重要的。来自中国多个机构的研究人员将地震记录和大地测量相结合，来综合评价华北地区的地震潜力。研究人员首先对地震目录进行了评估，并使用不同最小震级断点 (**cutoff**) 目录，通过震级-频率分布估计 b 值，确定了震级的完整性范围为 **1.9~2.8**。然后，利用全球导航卫星系统 (**GNSS**) 的震间观测数据，得到了三个应变速率参数 (最大剪切应变速率、主应变速率和应变速率第二不变量) 的空间分布。之后，基于应变速率和 b 值对华北地区的地震危险性进行了分析，确定了地震危险性较高的区域。结果表明，较大应变速率主要集中在鄂尔多斯地块西缘和唐山断裂带。不同的应变速率参数所产生地震重现期 (**recurrence interval**) 的空间分布比较相似，但大小不同。鄂尔多斯地块西缘和唐山断裂带与华北其他地区相比，呈现出较短 (**7000~15000** 年) 的 M_L 7.5 级地震重现期。本研究结果将有助于华北地区地震危险性的评估。相关成果发表在 **2023** 年 **12** 月的 ***Tectonophysics***。

华北位于亚洲大陆东部，由于板块碰撞，形成了许多断层。该地区经历了 24 次以上的破坏性地震 (图 1b)，由于众多大城市位于华北地区，这导致了相当大的生命和财产损失。以往的研究主要基于历史大地震评估地震危险性，然而，这些结果的有效性受到地震记录的完整性，特别是古地震资料完整性的限制。许多研究也利用运动学模型来评估华北地区的应变和断层滑动速率，并揭示它们与地震的关系。然而，这些结果主要是现今的地壳应变速率或所累积的断层滑动亏缺 (**deficit**)，而不能用于估计华北地区未来的地震危险性。

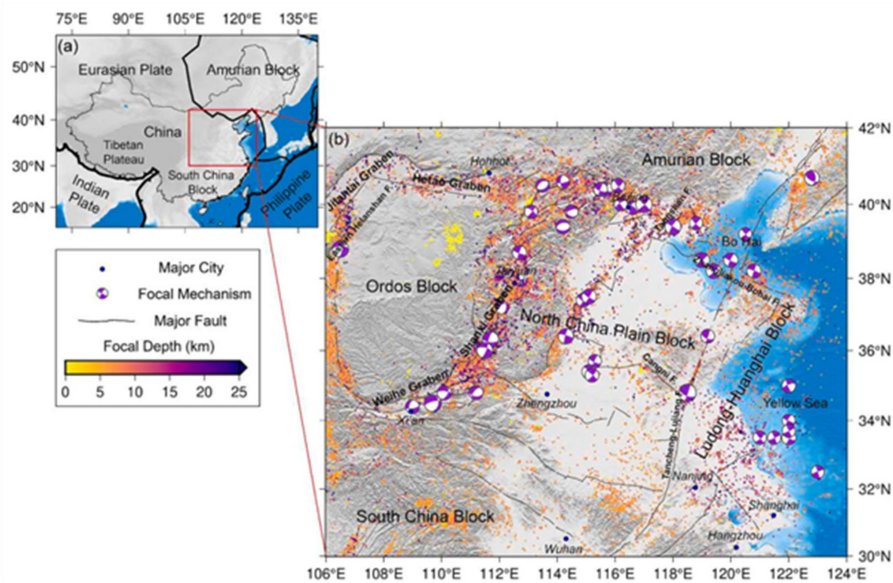


图 1 华北地区大地构造图。(a) 红色矩形代表研究区域；(b) 散点表示微地震，不同颜色表示不同震源深度

为了对华北地区地震危险性进行深入评估，研究人员首先利用 1995—2016 年全球导航卫星系统 (GNSS) 的震间观测数据，计算应变速率参数。然后，根据 2010—2021 年的微地震记录，使用古登堡-里克特定律估计了 b 值。之后，利用获得的应变速率和 b 值估计地震重现期，并对不同参数的影响进行详细评估。最后，对华北地区的地震危险性进行了分析，确定了地震危险性较高的区域。

1. b 值空间分布

b 值的空间变化与构造应力差异和地壳物质的非均质性有关，是地震危险性评价的重要指标和前兆。 b 值通常接近 1.0，较大的 b 值表明地震群较小，较低的 b 值表明大地震占主导地位。根据不同震级目录，研究人员得到了 b 值的空间分布和不确定性，在鄂尔多斯地块东北部、苍尼断裂和郟城-庐江断裂北段观测到较大的 b 值。通常 b 值随着最小震级断点 (cutoff) 的增加而增加，从 0.91 上升到 0.94。当使用最小震级标度 ($M_L=2.0$) 的断点时，鄂尔多斯地块东北部和郟城-庐江断裂北段的 b 值较大 (图 2b)。

在唐山断裂带周围发现较低的 b 值，这可能是由于地震目录中发生了 7 次以上 $M_L \geq 4.0$ 级地震。不确定性比 b 值低一个量级，表明了结果的合理性（图 2c 和 2d）。

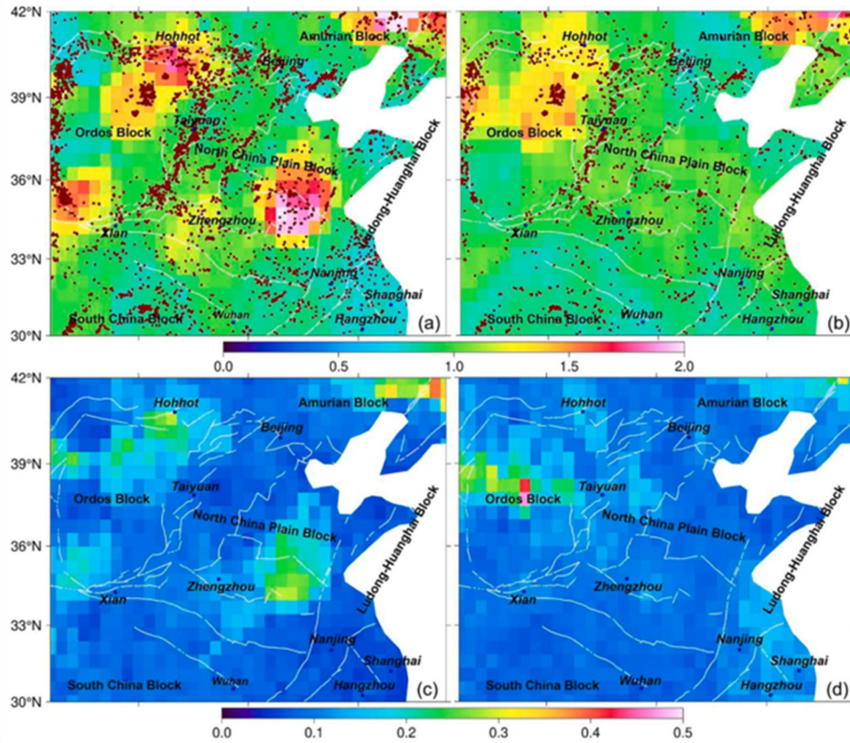


图 2 (a) 通过最小 $M_L=1.5$ 断点估计的 b 值的空间分布和不确定性 (c)，暗红点表示相应的地震，(b, d) 类似于 (a) 和 (c)，但最小 $M_L=2.0$ 断点

2. 应变速率

华北地块和鄂尔多斯地块内部的最大剪切应变速率 $\dot{\epsilon}_{\text{shear}}$ 小于 $3 \times 10^{-9}/\text{年}$ ，表明稳定和刚性的运动。相对较大的 $\dot{\epsilon}_{\text{shear}}$ 集中在鄂尔多斯地块西缘、张家口-渤海断裂、唐山断裂和山西地堑北端，大约为 $17 \times 10^{-9}/\text{年}$ ，表明强烈的剪切作用（图 3a）。华北地区主压缩应变速率以北北东-南南西和北东-南西向为主。唐山断裂带外围区域主拉伸应变速率较大（ $22 \times 10^{-9}/\text{年}$ ）。鄂尔多斯地块西域主压缩应变速率约为 $18 \times 10^{-9}/\text{年}$ （图 3b），说明该地区受到青藏高原强烈的推挤运动的控制。

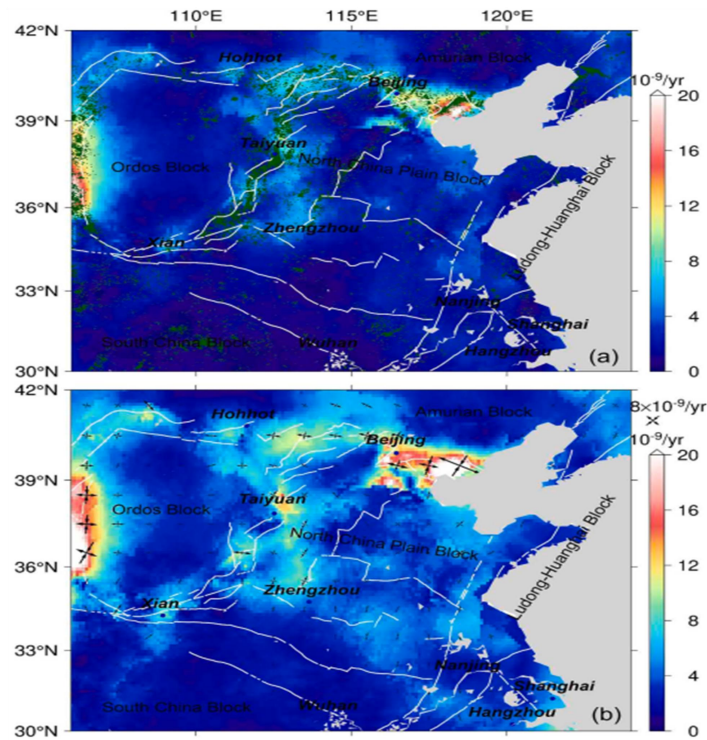


图3 使用 GNSS 观测估计的应变速率。(a) $\dot{\epsilon}_{\text{shear}}$; (b) 主应变速率张量和 $\dot{\epsilon}_{2\text{inv}}$ 的等值线

代表全应变速率场的应变速率第二不变量 $\dot{\epsilon}_{2\text{inv}}$ 在空间分布上与 $\dot{\epsilon}_{\text{shear}}$ 相似，但量级更大，特别是在山西地堑、唐山地区和鄂尔多斯地块西部（图 3b），表明这些区域受到剪切和伸展/挤压变形的双重控制。渭河地堑中段、河套地堑西部和山西地堑中段的 $\dot{\epsilon}_{\text{shear}}$ 达到 $5 \times 10^{-9}/\text{年}$ ， $\dot{\epsilon}_{2\text{inv}}$ 高达 $14 \times 10^{-9}/\text{年}$ ，表明这些地区以拉伸为主。总体而言，该研究计算的应变速率的空间分布和大小与前人构建的模型一致。

3. 地震活动与应变速率的相关性

为了评估应变速率是否可以用于估计华北地区地震的重现期，研究人员评估了应变速率与地震活动之间的关系。研究人员采用 3 个应变速率参数和 4 个地震目录进行相关性分析。在 4 个地震目录的初始阶段，所有应变速率参数都与地震数量有很强的相关性。而在曲线的中间部分，可以发现相对较大的差异，特别是 $\dot{\epsilon}_{\text{shear}}$ 和 $\dot{\epsilon}_{\text{max}}$ （最大主应变速率），这可能是由

于 GNSS 数据获得的应变速率分辨率较低所致。因此，在随后的分析中将使用这些不同的应变速率来估计地震重现期，并比较它们各自的贡献。

4. 讨论

(1) b 值和最大震级 M_{\max} 对地震重现期的影响

研究结果表明，在相同应变速率条件下， b 值越大，地震重现期越长，如鄂尔多斯地块西南边界与苍尼断裂带。此外，研究人员通过假设 $M_{\max}=7.8$ 和 8.2 来评估 M_{\max} 对重现期的影响。结果显示， M_{\max} 越大，重现期越长，但 M_{\max} 的选择并没有改变重现期的空间分布样式。尽管空间尺度上略有差异，鄂尔多斯地块西部与张家口-渤海断裂带的地震重现期相对较短，表明这两个地区具有较高的大地应变速率，且地震次数与应变速率之间具有良好的相关性，意味着未来可能存在较高的地震危险性。

(2) 应变速率对地震重现期的影响

研究人员进一步利用 $\dot{\epsilon}_{\text{shear}}$ 和 $\dot{\epsilon}_{2\text{inv}}$ 计算了 $M_L7.5$ 地震的重现期。利用 $\dot{\epsilon}_{\text{shear}}$ 所估计地震重现期的空间分布与利用 $\dot{\epsilon}_{\text{max}}$ 估计的地震重现期的空间分布相似。两个结果都表明，鄂尔多斯地块西部和唐山断裂带的地震重现期约为 1~2 万年。通常，主应变速率的极值大多大于 $\dot{\epsilon}_{\text{shear}}$ 。因此，用 $\dot{\epsilon}_{\text{shear}}$ 估计的地震重现期会更长。由于 $\dot{\epsilon}_{2\text{inv}}$ 大于 $\dot{\epsilon}_{\text{max}}$ 和 $\dot{\epsilon}_{\text{shear}}$ ，因此估计的地震重现期将是三个应变速率参数中最短的。

(3) 中国华北地区的地震危险性

研究人员根据大地应变参数和 b 值，获得了华北地区地震重现期的分布。然后，根据重现期进一步得到华北地区潜在地震的空间分布（图 4）。结果显示，鄂尔多斯地块西缘和唐山断裂带的地震重现期相对较短，这与前人的结论一致。这两个地区都显示出大约 7000~15000 年的 $M_L7.5$ 地震重现期。

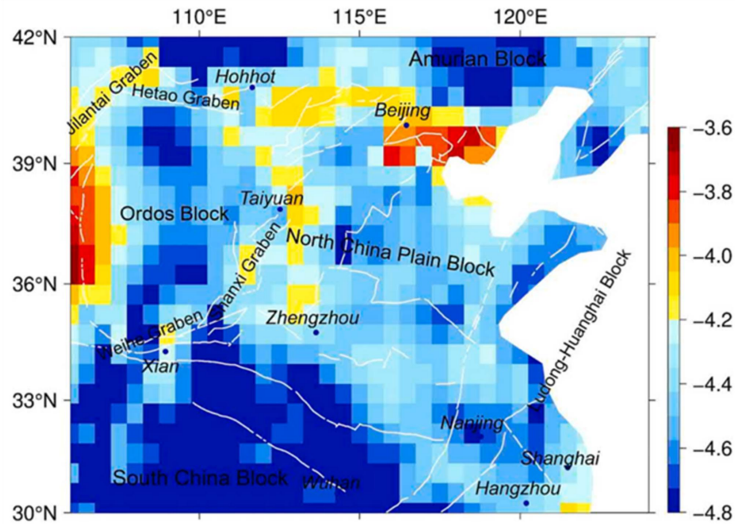


图4 利用最小 $M_L=2.0$ 断点和 $M_{max}=8.0$ 估计的 ϵ_{2inv} 和 b 值来评估华北地区 $M_L=7.5$ 地震的可能性，以 \log_{10} 速率/年显示

1978 年发生在唐山断裂带上的唐山 7.8 级地震是最近发生在华北平原的一次破坏性地震，1739 年发生在鄂尔多斯地块西部的平罗 8.0 级地震是有记录以来最大的板内正常断裂地震之一。青藏高原东北向运动对鄂尔多斯地块产生强烈的推挤作用，进一步导致了鄂尔多斯地块西南边缘的压缩变形和顺时针旋转，并在其西部边缘产生了右向剪切作用。基于地球物理和大地测量的模拟表明，鄂尔多斯地块西缘也存在较高的累积应力。在唐山地区，GNSS 观测显示华北平原北部存在剧烈的变形梯度。此外，地球物理研究表明，薄岩石圈结构可能归因于太平洋板块俯冲引起的热上升流。同时，该区域具有低地震速度和低热流的特征。

此外，引起华北地区地震活动的另一个因素是地下水的卸载和地面沉降，这可能导致应力调节和地震。但是，鄂尔多斯地块西缘与唐山地区地面沉降较少。因此，这些地区的地震活动受水文加卸载影响的可能性较小。

来源：Gao Y, Qu W, Zhang Q, *et al.* Assessment of the seismic hazard in North China by combining micro-seismicity records and geodetic observations. *Tectonophysics*, 2023, 869: 230130.

利用大数据和人工智能进行地震预报：一项为期 30 周的中国实时案例研究

地震预报旨在拯救人类生命并减轻灾难性损失，是地震学领域最具挑战性的任务之一。在本研究中，研究人员设计了一个实时地震预报框架来预报地震，并在中国西南地震区进行了测试。输入数据是多分量地震监测系统 AETA（声、电磁到人工智能）提供的特征，其中每个台站使用两种类型的传感器记录数据：电磁（**electromagnetic, EM**）传感器和地声（**geo-acoustic, GA**）传感器。研究目标是根据当周的数据预测下周可能发生的地震的位置和震级。所建议方法以使用主成分分析对大量 EM 和 GA 数据进行降维为基础，然后进行基于随机森林的分类。所建议算法使用 2016—2020 年的可用数据进行训练，并使用 2021 年的实时数据进行评估。结果表明，测试准确率达到 70%，而精确度、召回率分别为 63.63%、93.33%。与目录解相比，使用该方法预测的距离和震级的平均绝对误差分别为 381 km 和 0.49 级。相关研究成果发表在 2023 年 12 月的 *Bulletin of the Seismological Society of America*。

地震是一种自然灾害，可能对人类生命、基础设施、环境、财产和建筑物造成严重损害。地震是构造活动的结果，且可能不是周期性的。地震发生突然等特征使地震预报工作具有挑战性，地震预报系统的目的是预报地震及其震源参数，即位置和震级。地震预报可以是短期或长期发布，当预报系统发出准确的警报时，可为预防和准备行动提供支撑，以减少可能的损失。这对容易发生地震灾害的中国来说意义重大。前兆信号的探测是地震预报的关键之一，其通过多学科方法来开展观测，而人工智能（AI）技术以前所未有的方式为检测地震前兆信号打开了大门。

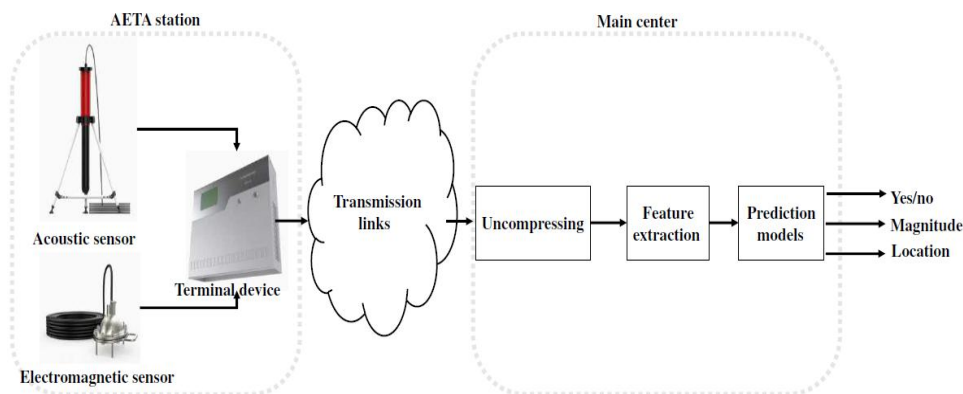


图 1 AETA 台站的组成。每个台站有两个传感器：声学传感器和电磁传感器，终端设备用于收集两个传感器的数据

本研究主要报告了 2021 年国际 AETA 地震预报大赛中第一个基于人工智能的实时地震预报案例。该项目的目标是利用当周的 AETA 特征来预测下周的地震。如果发生 3.5 级以上的地震，则应根据相同的提取特征来预测地震的位置和震级。在研究方法上，基于奇异值分解（SVD）使用主成分分析（PCA）进行降维。然后，基于 SVD 值，使用随机森林（RF）来预测地震的发生。在这种情况下，地震预报是一个二元分类问题，其中使用两个类别：0 类表示无地震，1 类表示有地震。此外，在下周可能发生地震的情况下，利用 PCA 和 RF 方法来预测地震的位置和震级。

1. 输入特征分析

首先，研究人员分析了对增强结果影响最大的重要特征。GA 数据最重要的特征是偏度、均值、重心频率、峰度以及小波变换等。对于 EM 数据，重要的特征是峰度、超低频峰度、超低频功率和偏度等。在本研究中，前兆信号被认为是 EM 和 GA 数据中具有一定幅度和模式的扰动。前兆信号的幅度可以从时域检测，而前兆模式在频域和时频域有更好的表示。RF 模型强调的一些最重要的特征是从时域中提取的，该时域表示前兆信号的

幅度，例如偏度、均值和峰度。另一方面，一些特征突出了前兆信号在频域和时频域中的模式，例如中心频率、超低频和多级小波变换等。

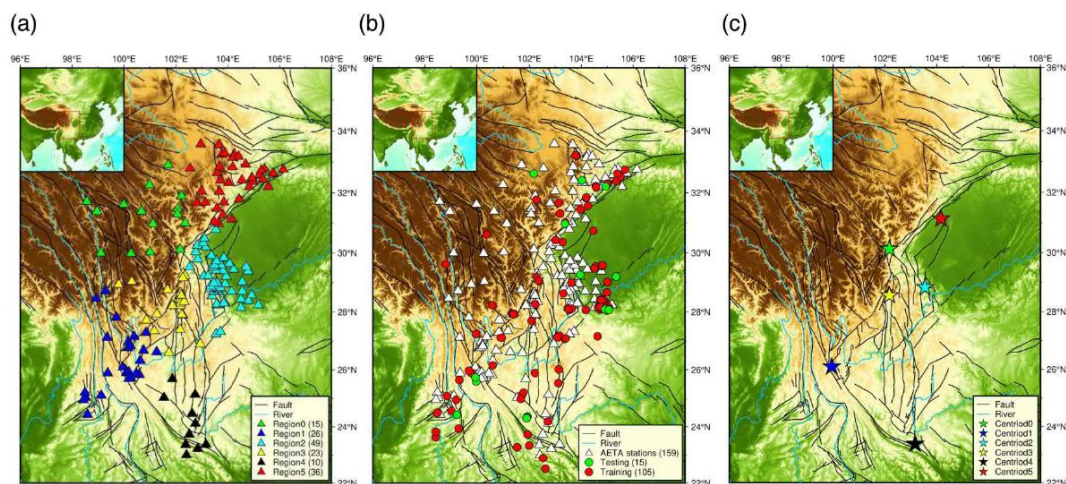


图2 (a)AETA 站的分布，(b)训练集和测试集的位置，(c)六个 K-均值簇的质心。插图表示研究区域（红框）

2. EM 和 GA 数据及簇数量的影响

有了前述基础之后，研究人员评估了 EM 和 GA 数据对地震预测的影响。最开始，只使用 GA 数据并忽略 EM 数据，即每个站的输出概率（output probability，分两类，即地震类、无震类）仅基于 GA 传感器。然后，使用 EM 数据进行相同的测试，并忽略 GA 数据。结果表明，仅使用 EM 数据会降低所提出框架的准确性，而使用 GA 数据与同时使用 EM 和 GA 数据具有几乎相同的性能。这说明研究中使用的特征选择策略很可能偏向 GA 信号。事实上，电磁信号对地震预报的可靠性仍然存在一定争议。

研究人员进一步评估了将目标区域划分为不同数量子区域时的影响。根据站点的位置将目标区域分为 4、6、8 和 12 个簇，并使用相同的预测策略。当簇数量增加时，选择正确区域（距地震最近）变得更加困难。研究人员选择概率最高的子区域中心作为预测的地震位置。结果表明，8 簇

情况的精确度与 6 簇情况相同，但真阳性 (TP) 比率较低，距离误差较大。同时，6 簇情况时，距离误差最低，精确度误差和震级误差合理。

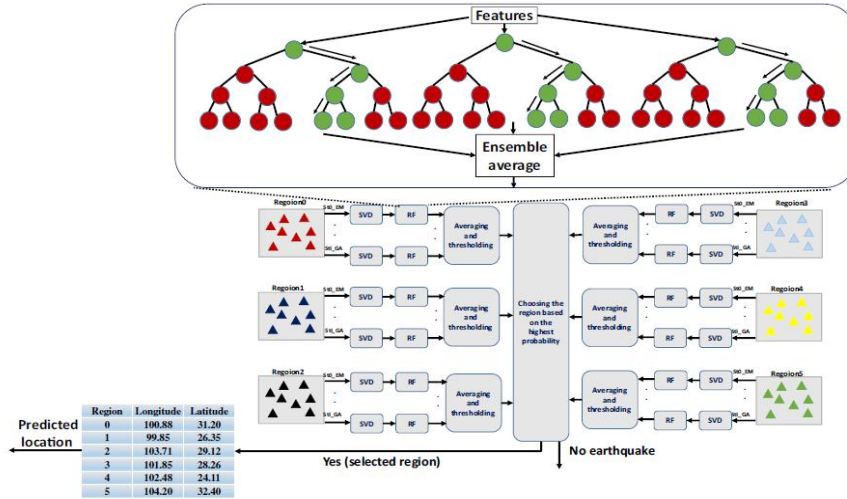


图 3 使用 AETA 数据进行中国地震预报的拟议方法框架。RF 指预测地震发生（有地震或无地震）的随机森林模型，Sti 表示每个区域的站点数量

3. 平衡数据集，选择输出阈值

为平衡训练数据集，研究人员使用了合成少数过采样技术 (SMOTE)。具体来说，SMOTE 生成与地震类 (earthquake class) 相似的合成数据 (SVD 数据)，使得每个类的训练数据集规模一致。研究人员使用相同的策略来平衡震级估计模型的数据集，结果表明，所提方法对大地震没有显著偏差。然而，未来的研究需要进一步增强从 EM 和 GA 数据中提取的特征。

为了评估所提出算法的可靠性，将目标震级更改为 $M_L 5$ 。结果表明，所提出的算法在使用不同目标震级时具有稳健的性能。由于所提出算法将目标区域分为六个区域，每个簇都有自己的输出阈值。当 6 个区域的阈值增加时，精确度和召回率均下降。另一方面，当降低阈值时，相应地精确度下降，召回率增加。因此，需要选择合适的阈值以实现所提出算法的最佳性能。对于 6 个区域而言，这一数值为 0.6~0.8。

研究所提出的数据驱动的机器学习方法是一种短期模型，可以预测每周或每天地震的发生。尽管预报精度不高，且空间分辨率相对较小，但该方法可以补充现有的长期或中期模型，以更高的时间分辨率预报地震。该方法预测的震中可以为区域预测模型添加新的约束，以获得更高的空间分辨率。

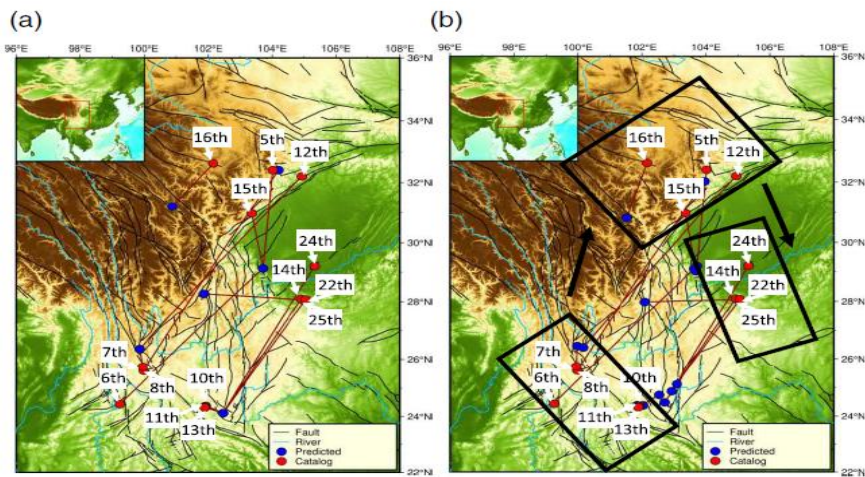


图4 与目录位置相比下的地震预测位置。(a)预测位置被视为每个聚类区域的质心，(b)考虑上月地震位置的情况下，定位精度得到提高。白色标签表示目录地震的星期数，标记的目录地震揭示了清晰的地震活动偏移模式

4. 小结

中国是世界上地震最活跃的地区之一。在本研究中，研究人员使用AETA提供的EM和GA数据来预测中国的地震。每周有1008个10分钟的窗口，每个窗口包含44个GA数据特征和51个EM数据特征。PCA（主成分分析）被用于减少RF模型的输入数据大小，以预报下周可能发生的地震。在没有地震的情况下，这些模型的输出为0（无震），在发生地震的情况下，这些模型的输出为1（有震）。

根据站点的位置，研究区域被分为六个子区域。然后，对于每个区域，根据属于该区域的台站来预测有地震或无地震。研究人员对同一区域内不同站点的输出预测进行平均，这样就有了对应于六个区域的六个输出概率。

之后，进一步为每个区域设置一个阈值。如果区域的输出概率超过相应的阈值，就认为预测结果是 TP（有地震类），否则认为它是 TN（无地震类）。

由于我们将原始空间缩减为子区域，因此我们将与其阈值相比概率最高的区域中心视为地震的最终可能位置。对于震级估计，研究人员为网络中的每个传感器设计了一个 RF 模型，其中的输入与地震预测中使用的奇异值相同。震级被分为五类，其中最终震级是预测类的质心。该算法的结果令人鼓舞，因为在 30 周中正确预测了 21 周，准确率达到 70%。同时结果还表明，影响稳健性的最重要特征是峰度、偏度、频率成分和小波级别。然而，可以使用基于物理的约束来增强当前的框架，例如区域的速度模型、断层分布。未来，所提出的预测方法可以在不同的地理区域进行测试，包括地震活动较弱的区域。

来源：Saad O M, Chen Y, Savvaidis A, *et al.* Earthquake Forecasting Using Big Data and Artificial Intelligence: A 30-Week Real-Time Case Study in China. *Bulletin of the Seismological Society of America*, 2023, 113(6): 2461-2478.

利用梯度提升回归器根据 P 波特征实时预测距离和 PGA 在现地预警中的应用

现地预警（on-site EEW）系统是减少地震灾害的重要手段。来自意大利的研究人员使用梯度提升回归器（Gradient Boosting Regressor, GBR）根据 P 波特征来预测峰值地面加速度（PGA）以及震中距离（D）。研究人员使用意大利中部 Amatrice-Norcia-Visso 地震序列的波形数据集（2016—2017 年）来验证和训练模型，并在拉奎拉地震序列（2009 年）的数据集上对模型进行了测试。从预测得分的角度研究了 GBR 预测 D 和 PGA 的性能，发现即使使用 1 s 窗口，模型也可以很好地预测两个目标。研究人员提出了一个基于 D 和 PGA 预测的概率现地预警系统模型，并测试了两组不同的阈值：感知预警（felt alert）和破坏预警（damage alert）。同时，研究人员利用 PGA 的提前期（lead time, LT）来区分有用的预警（积极的 LT）和错过的预警（MA）。最后，研究人员分析了这种系统的性能，考虑了四种可能的情况：成功预警（SA）、错过预警（MA）、高估预警（OA）和低估预警（UA）。结果表明，在 1 s 窗口时系统的 SA 率约为 80%，随着 MA 的增加，系统的 SA 率降至 65% 左右。这一结果表明，所提出的系统在 1 s 窗口的情况下已经是可靠的，这对于意大利中部这样的地震易发地区（具有中到大地震风险）来说具有巨大优势。相关研究成果发表在 2024 年 1 月的 *Geophysical Journal International*。

地震对人类生命、财产安全产生了巨大威胁。如今，实时地震学通过减少人员和目标对地震潜在破坏性影响的暴露度，成为减轻地震风险的最有效措施之一。此外，自动学习算法的发展，如人工智能和机器学习，可以优化从数据中提取信息的过程。在过去的几十年里，地震学领域的许多学者都提出了利用机器学习和深度学习来实时研究地震的新技术。

研究人员探索了使用机器学习回归实时预测震中距离（D）和峰值地面加速度（PGA）的可行性。具体而言，使用了意大利中部发生的两个地震序列的波形大数据集，以 2016—2017 年中意大利序列为主的地震序列 S1619 被用于模型的训练（S1619_T）和验证（S1619_V），2009 年的拉奎拉序列被用于模型的测试（S09）。研究人员推导并讨论了将 P 波信息与距离和地面运动强度（指 PGA）联系起来的模型的性能，还在一个原型概率警报系统的框架内测试了现场预测。

1. GBR 用于震中距离的预测

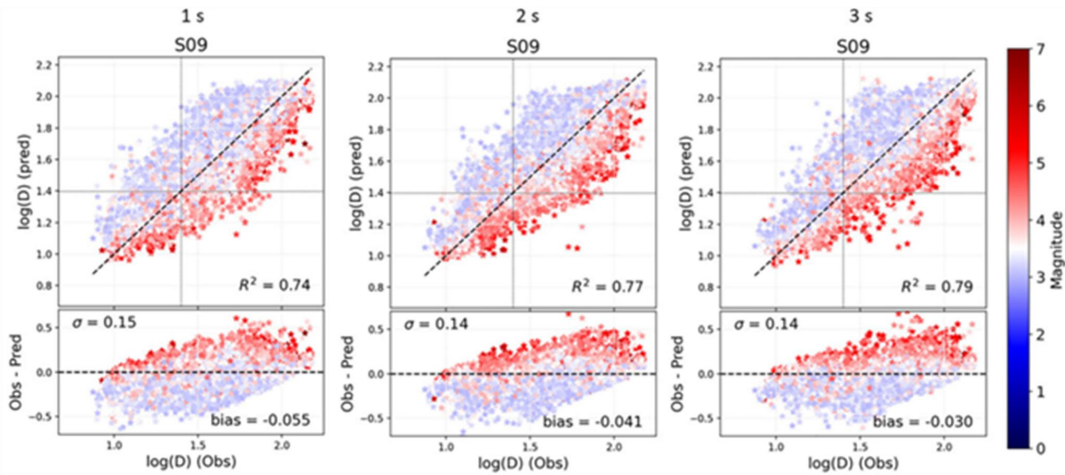


图 1 基于 S09 数据集在三个时间窗所预测 $\log_{10}D$ 与观测值的关系

图 1 显示了基于 S09 数据集在 1 s、2 s 和 3 s 窗口时对震中距离对数（ $\log_{10}D$ ）的预测。数据按震级着色，以突出震级和距离之间的平衡。结果表明，预测是可靠的，在 1 s、2 s 和 3 s 的时间窗口分别有 $R^2=0.74$ 和 $\sigma=0.15$ ， $R^2=0.77$ 和 $\sigma=0.14$ 以及 $R^2 = 0.79$ 和 $\sigma=0.14$ 。正如预期的那样，这些值低于验证结果，但它们仍然表明了良好的预测性能。此外，随着时间窗长度的增加，结果略有改善。

残差面板（图 1 中较低的子图）很好地突出了残差与震级的强烈依赖性。结果证实，任何旨在通过 P 波预测震级的现地预警方法都很容易存在很

大的不确定性。为了更好地评估模型的性能，研究人员分析了接收者操作特征（Receiver Operating Characteristic, ROC）曲线。从三个数据集的 ROC 曲线来看，训练集（S1619_T）的表现最好，而验证集（S1619_V）和测试集（S09）的结果略差。尽管存在过拟合，S09 数据集在所有时间窗口中的 AUC（指 ROC 曲线下方的区域）的值都大于 0.92，这代表了特别好的性能。

2. GBR 用于 PGA 预测

图 2 展示了通过 GBR 获得的 $\log_{10}PGA_H$ 的预测，以及 S09 数据集在 1 s、2 s、3 s 窗口的残差情况。对于 1 s 窗口，结果已经非常好了， R^2 值等于 0.90。对于较长的时间窗，可以观察到预测的轻微改善，在 3 s 时 R^2 达到 0.95，标准差范围从 1 s 窗口的 0.27 降到 3 s 窗口的 0.19。与之前一样，研究人员计算了 S1619_T、S1619_V 和 S09 数据集的 $\log_{10}PGA_H$ 预测值的 ROC 曲线和 AUC 值。同样，在这种情况下，可以观察到一个小的过拟合。尽管如此，GBR 模型在预测 PGA_H （水平方向的 PGA）方面表现出了非常好的性能。

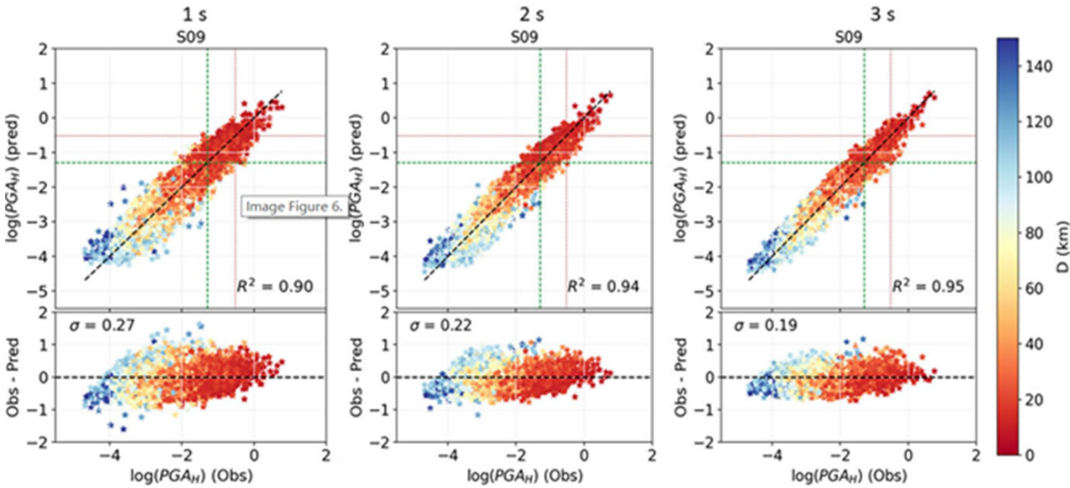


图 2 基于 S09 数据集在三个时间窗所预测 $\log_{10}PGA_H$ 与观测值的关系

3. 地震预警系统原型

为了强调系统的潜在优势和局限性，研究人员建立了一个单站预警系统原型。在这个框架中，即将到来的地震破坏的预警通知是基于对 P 波地面运动的局部测量而发出的，不需要对震源参数进行精确估计。研究人员定义了两个级别的预警：破坏预警（DA）和感知预警（FA）。此外，定义了 PGA_H 中与这些预警级别相关的阈值，FA 时的强度为 IV，DA 时的强度为 VI。FA 中 D 的阈值定为 50 km，DA 中则为 25 km。

于是，获得四个不同级别的预警：①警报 0：距离>阈值， $PGA <$ 阈值（没有危险）；②警报 1：距离<阈值， $PGA <$ 阈值（很近，但是危险可忽略）；③警报 2：距离>阈值， $PGA >$ 阈值（较远，但具有危险）；④警报 3：距离<阈值， $PGA >$ 阈值（距离近且有危险）。

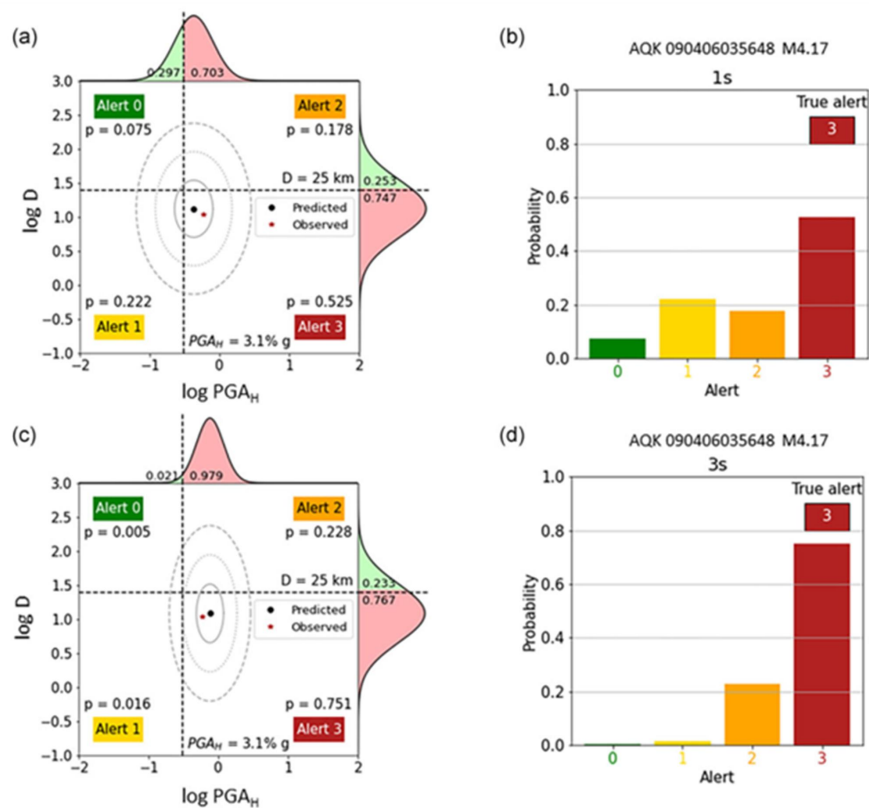


图 3 基于概率阈值的预警系统在站点 AQK 所记录 M4.17 地震中的应用

图 3 显示了 AQR 台站记录的震级为 M_w 4.17 的地震,并考虑了 DA 阈值。观测到的震中距离 D 和 PGA_H 分别为 11.0 km 和 0.580 m/s^2 (5.92% g)。因此,根据决策框架,预警系统的正确输出应该是警报 3。根据与 1 s 时间窗相关的信息,GBR 模型预测 D 等于 13.2 km, PGA_H 等于 0.426 m/s^2 (4.35% g)。这些预测值与实际值的误差在 1σ 以内 (图 3a)。这些对 D 和 PGA_H 的估计使预警框架以 53% 的概率正确地将预警级别输出为警报 3 (图 3b)。

图 3c 和 3d 给出了相同地震的预测结果,但使用了 3 s 窗口。对于这个时间窗长度,模型预测 $D = 12.2 \text{ km}$, $PGA_H = 0.756 \text{ m/s}^2$ (7.71% g)。因此,同样在这种情况下,即使考虑到这种情况下的标准差较低,目标的预测值与真实值的距离也在 1σ 以内 (图 3c)。地震预警框架再次提供了正确的预警级别,这次的概率为 75% (图 3d)。

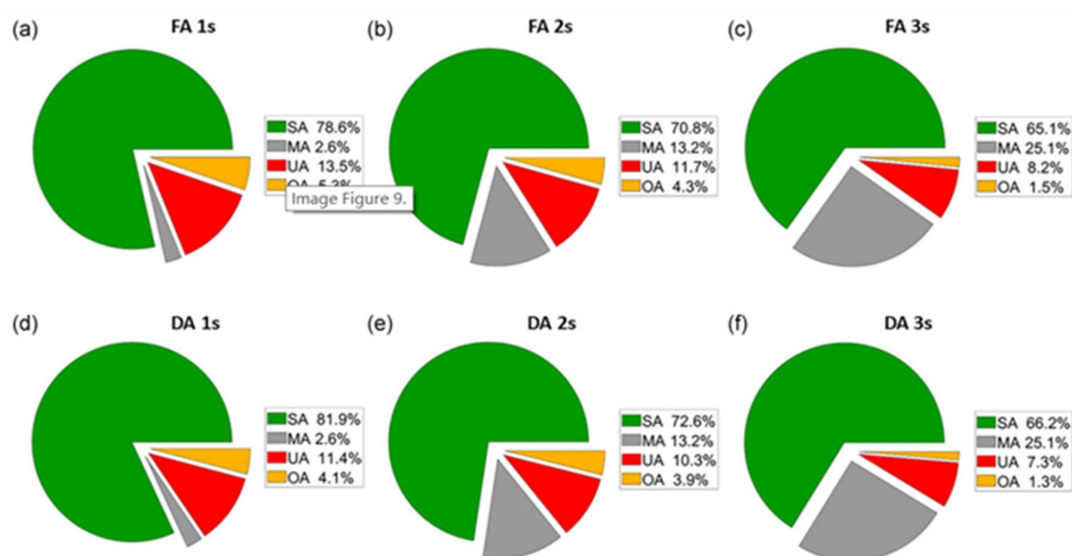


图 4 地震预警系统在 S09 数据集上的性能

图 4 显示了考虑每个时间窗长度和阈值设置时,整个 S09 数据集的地震预警性能。研究人员将 GBR 预测的 D 和 PGA_H 与观察到的实测值进行比较,以评估其性能。为了评估地震预警系统的性能,还必须考虑发生在离站点/目标太近而无法释放有用警报的地震。最后,为了简化问题,研究

人员定义了 4 种类型的表现：成功预警 SA、错过预警 MA、低估预警 UA 和高估预警 OA。定义如下：①SA：发出的预警=真实预警；②MA： $LT \leq 0$ s；③UA：发出的预警<真实预警；④OA：发出的预警>真实预警。

结果显示，GBR 的预测总是获得超过 65% 的 SA 率，在 1 s 窗口时 DA 的最大值为 81.9%。事实上，SA 随着时间窗口的增加而减少，这是由于 MA 的增加（即大窗口增加了盲区大小），后者的范围从 1 s 时的 2.6% 增加到 3 s 时的 25.1%。同样，在更大的时间窗下，UA 和 OA 也更低，而在这种情况下，MA 也会增加。这些结果表明，1 s 窗口的 GBR 模型在意大利中部容易发生中大型地震（ $M_w < 7$ ）的地区具有很好的地震预警潜力。将时间窗口增加到 2 s 或 3 s 将略微提高目标预测的精度，但代价是缩短 LT。此外，没有观察到 FA 和 DA 阈值之间的性能差异，这证实了所提供模型的稳健性和可靠性。

来源：Iaccarino A, Cristofaro A, Picozzi M, *et al.* Real-time prediction of distance and PGA from P-wave features using Gradient Boosting Regressor for on-site earthquake early warning applications. *Geophysical Journal International*, 2024, 236: 675–687.