

年变提取及预测评估软件

EQSuite-EVA V1.0

一、简介

时序数据中的破年变异常在地震前具有一定普遍性,针对该类现象本软件主要基于 **S** 变换时频方法构建了新的信息提取流程,相较于常规方法,一方面可避免正常年变背景拟合所碰到的问题,另一方面除了提取出常规短周期破年变信号 (**ONA**) 外还可以同时分析背景年变信号 (**ANA**) 的演化过程。在此基础上,基于双向非对称阈值策略,结合 **R** 值评分及 **Molchan** 图表法构建了预测指标确定和效能评估的一套定量方法,由于阈值参数空间的拓宽和更多评估参数的引入,使得指标选取和效能评估更为合理。

研发平台: Windows10/Matlab

面向行业/领域: 地震预测

最后更新时间: 2022-3-1

作者: 刘琦

Email: liuqi@ief.ac.cn

二、核心模型方法

1. 基于时频分析方法的破年变信息提取流程

考虑破年变异常的定义、特征（正常状态下观测数据应该具备相对规律的年周期信号背景，当年周期信号不再规律或者叠加一些短周期的畸变信号后，认为状态不正常，出现异常）以及时频分析方法在研究不同信号组分的非稳态时频演化过程的优势，基于时频分析方法可以实现破年变信息的分离和提取。

我们选择应用相对广泛的 **S** 变换方法作为处理流程构建的基础，该方法具有相对较好的时频聚集性和时频分辨率。

第一步：数据预处理。对于采样间隔小于等于天尺度的数据，需要进行去台阶、去突跳点、缺数插值等预处理，并降采样为日值数据用于后续计算。对于跨断层观测等采样间隔较大的数据，则需要用缺数标记将原始时序填补成等间隔的日值观测序列，之后利用样条函数进行插值。需要注意的是，原始观测数据采样间隔如果相对较大或者不规律，则可能会影响后续处理效果。

第二步：数据筛选。当数据观测时长足够并且观测数据具备相对清晰的年变背景时，开展破年变分析才有意义，因此在数据观测时长筛选的基础上，进一步通过对比傅里叶谱中年频段与其它频段幅度比，挑选年变清晰的数据用于后续处理。这里给定的筛选条件为数据长度 ≥ 2 年，年频段最大幅值/其它频段最大幅值 ≥ 1 ，考虑到年变信号周期存在一定变化范围，因此这里给定年频段为 250~500 天，其它频段为周期 250 天以下，不考虑周期大于 500 天的低频趋势项。

第三步：**S** 变换处理。对筛选出来的经过预处理的数据进行 **S** 变

换，为避免边界效应，在 **S** 变换之前需要对数据双侧扩边，计算完成后对 **S** 变换结果双侧截边，其中单侧扩边的长度这里设定为总数据长度的 7.5%。为提高计算效率，对于较高采样率的定点前兆观测数据，在 **S** 变换时将计算周期定为 14~500 天，即不考虑周期较长的低频趋势项以及周期相对较短的信号成分，对于以月观测为主的跨断层观测数据，考虑其周期分辨率，计算周期定为 90~500 天。

第四步：信息指标建立。基于 **S** 变换结果建立两个信息指标，一个是年频段归一化平均幅度，简称 **ANA**，一个是其它频段归一化平均幅度，简称 **ONA**。**ONA** 与常规方法提取的短周期破年变残差信号具有一定类比性。**ANA** 反映的是背景年变信号的演化过程。无论 **ANA** 还是 **ONA**，两个信息指标都经过了归一化，因此无量纲。经过上述处理，观测数据中的背景年变演化和短周期变化等过程均可通过两个信息指标被定量描述。

2. 综合 R 值和 Molchan 图表的预测指标确定和效能评估方法

针对上述两个信息指标,开展预测指标确定和效能评估。基于数据开展地震预测首先需要给出异常定义,即区分正常变化和异常变化的标准,一般是给出异常阈值。针对异常阈值的定量确定主要基于混淆矩阵,考虑有震报准率、无震报准率、漏报率、虚报率等因素,通过不同因素的组合和折中来确定最优的异常阈值,给出相应的效能评估结果。其中应用较为广泛的 R 值评分就基于上述因素构建,由于各因素并非完全独立,因此 R 值评分有多种等价的表达形式,目前较为常用的一种形式主要考虑有震报准率和预报占时率。

R 值为扣除了随机概率后的预报成功率,当 $R=0$ 时表示预报没有起作用。一般除了考虑 R 值大小外,还需要考虑 R 值的信度。同样的 R 值,当其基于的地震次数不一样时,信度也不一样,因此还需要考虑 R_0 值,即保证 97.5%置信度的最低 R 值,当 R 值大于 R_0 值时可以认为这个 R 值至少有 97.5%置信度。 R_0 值仅与地震总次数和报对地震次数有关,当地震总次数较少时, R_0 值可基于二项式分布计算,当地震总次数相对较多时(例如大于等于 20 次), R_0 值则可基于正态分布计算。在地震总次数相同的情况下,基于两种分布获得的 R_0 值随报对地震次数变化的特征基本一致,其中绝大多数情况下,基于正态分布计算得到的 R_0 值相对较小。

此外, Molchan 图表法同样也是考虑漏报率、预报占时率的一种方法,并且可以通过概率增益、显著性水平等进一步针对模型预测能力开展有效评价(图 1a)。当给定了研究需要的震级范围、空间范围后,在研究时段内用于分析的地震总样本便被确定。进一步给定预报时窗 DT(即异常出现后预测多长时间内会在目标区域发生目标震级

的地震), 通过设置异常阈值, 可以获得在该阈值条件下异常的预报占时率以及相应的漏报率, 对应 **Molchan** 图表中的一个坐标点。动态设置一系列异常阈值, 则其相应的统计结果在 **Molchan** 图表中形成一条曲线。因为漏报率与报准率之和等于 1, 因此该曲线上下翻转后等价于以报准率为纵坐标的 **ROC** 曲线。基于对该曲线上各点的效能评估可以确定在特定规则下(特定的震级范围、预报范围以及预报时窗)的最优阈值。我们主要考虑距离对角线最远的点 **P1** 以及距离原点最近的点 **P2**。根据相关几何公式可以很容易推导出 **Molchan** 曲线上的点到对角线的距离等于 $R/\sqrt{2}$, 因此 **P1** 点实际上对应 **R** 值最大的点(即漏报率、预报率占时率的 **L1** 范数最小点), **P2** 点则考虑了漏报率、预报占时率的 **L2** 范数最小。当 **P1** 点和 **P2** 点不相同, 进一步从中选择概率增益更大的点(概率增益 $\text{Gain} = \text{有震报准率} / \text{预报占时率}$, 反映了相比随机概率, 目前概率提升的倍数)。同时, 基于显著性水平参考曲线可以相对方便的判断 **Molchan** 曲线点的信度情况, 例如当 **Molchan** 曲线点位于显著性水平 α 为 2.5% 的参考曲线左侧时, 表示该点显著性水平低于 2.5%, 其对应 **R** 值大于 **R0** 值(因为 **R0** 值对应的是置信度 97.5% 即显著性水平 2.5% 的情况)。

在阈值设置上, 一般做法是基于双向对称阈值(以零值线或均值线为对称轴)或单向阈值策略设置一系列阈值参数, 本质上阈值参数都是一维的。考虑到待检验的信息指标时序其正向和负向变化的机制以及对于地震的指示作用可能存在差异, 因此我们采用双向非对称阈值策略(即信息指标小于负向阈值 **Th1** 或大于正向阈值 **Th2** 均被判定为异常, $\text{Th2} \geq \text{Th1}$), 充分考虑正向和负向的差异性, 将阈值参数拓展为二维。通过这种处理方式, 同一漏报率可能会对应多个预报占时率, 可以选择预报占时率最小所对应的阈值参数组合。由于双

向对称阈值、单向阈值的情况实际均被包含在内，因此基于双向非对称阈值策略获得的阈值参数，其预测效能为三者中最优（图 1b）。

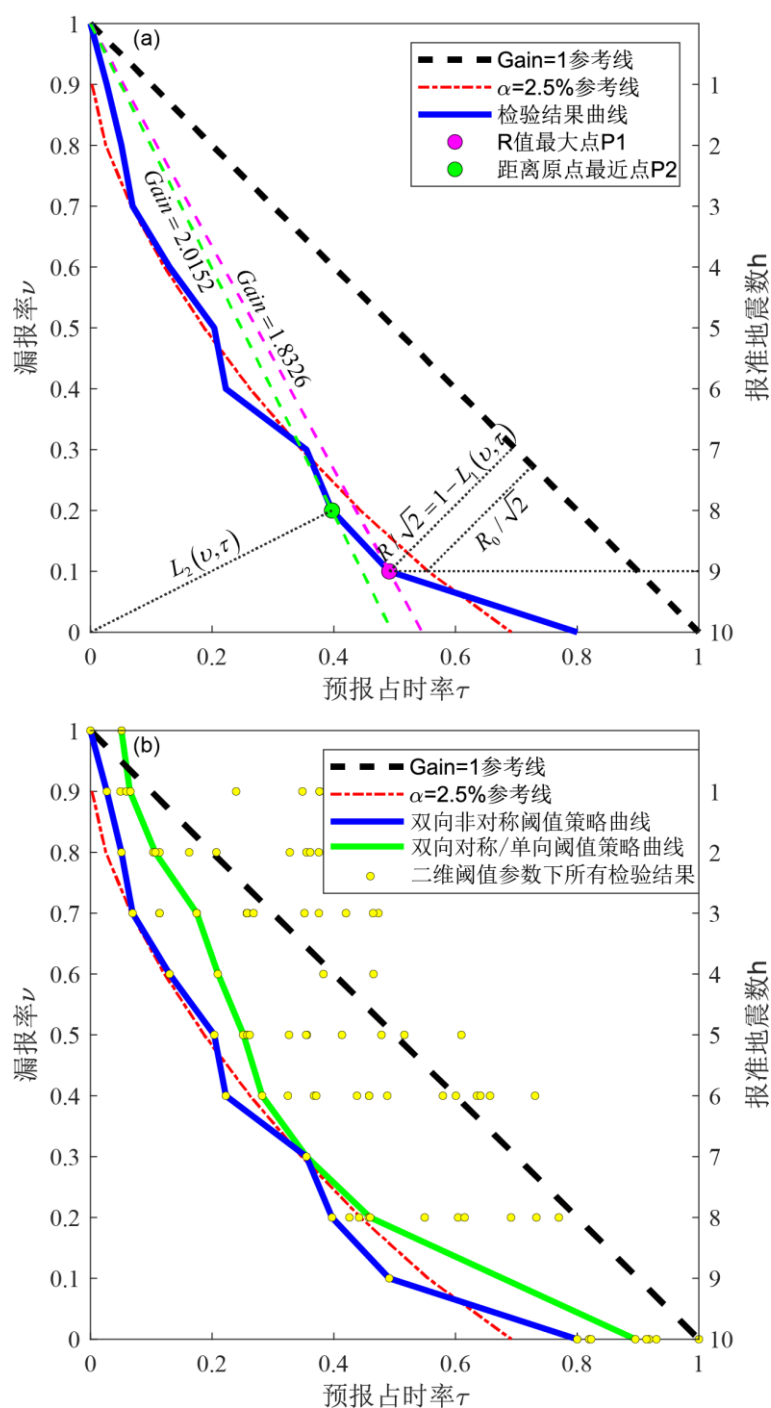


图 1 预测指标确定和效能评估定量方法的原理示意图

基于上述处理，可以获得在给定震级范围、空间范围、预报时窗下的相对最优阈值参数组合及相应的预测效能结果。将震级分档后，通过对空间范围、预报时窗等参数进行调整，重复相对最优参数指标的计算过程，可以从中确定最优的参数指标。具体做法为，在同一震级档下，对比不同空间范围、预报时窗下其相对最优阈值参数组合所

对应的 **R** 值，选择 **R** 值最大的空间范围、预报时窗、阈值参数作为该震级档的最优预测指标。当最大 **R** 值对应多个空间范围和预报时窗的参数组合时，可进一步通过概率增益、显著性水平等进行筛选，选择概率增益大、显著性水平低的参数组合。

三、实际数据应用

准备好 EQT 格式的地震目录文件以及索引文件，索引文件中每行依次给出台站经度、纬度、数据文件完整路径，以空格作为分隔符。在 Matlab 环境下点击运行 Main_ST_ANN_evaluationNew.m 即可。

这里以新疆库尔勒台水平摆倾斜北南测项为例，介绍软件的实际使用过程。使用的数据时段为 2008 年 1 月 1 日至 2022 年 2 月 22 日，期间台站周边中强地震相对较多，具备较好的研究基础。索引文件中的条目为“86.2 41.8 测试数据\65015_2_2211.Dat”。

程序运行后首先给出经数据格式转换、预处理、降采样、长度筛选后的年变显著程度检测结果，该数据的幅值比为 2.02，具备相对清晰的年变背景，满足继续计算的要求（图 2）。

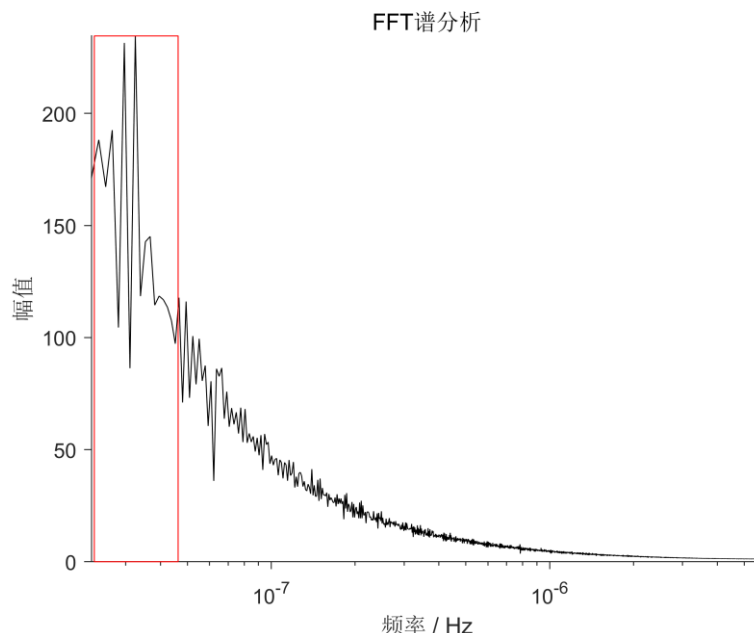


图 2 年变显著程度检测结果

基于 S 变换方法提取获得了年频段归一化平均幅度（ANA）和其它频段归一化平均幅度（ONA）两个信息指标（图 3），其中 ANA 反映了背景年变信号的演化过程，可以看到年变幅度缓慢增大并在 2012 年 7 月前后到达峰值，之后又开始减小，直到 2016 年 8 月前后

再次开始缓慢增大；ONA 显示了短周期信号的变化情况，在 2015 年之前变化相对剧烈，2015 年之后变化相对减缓。

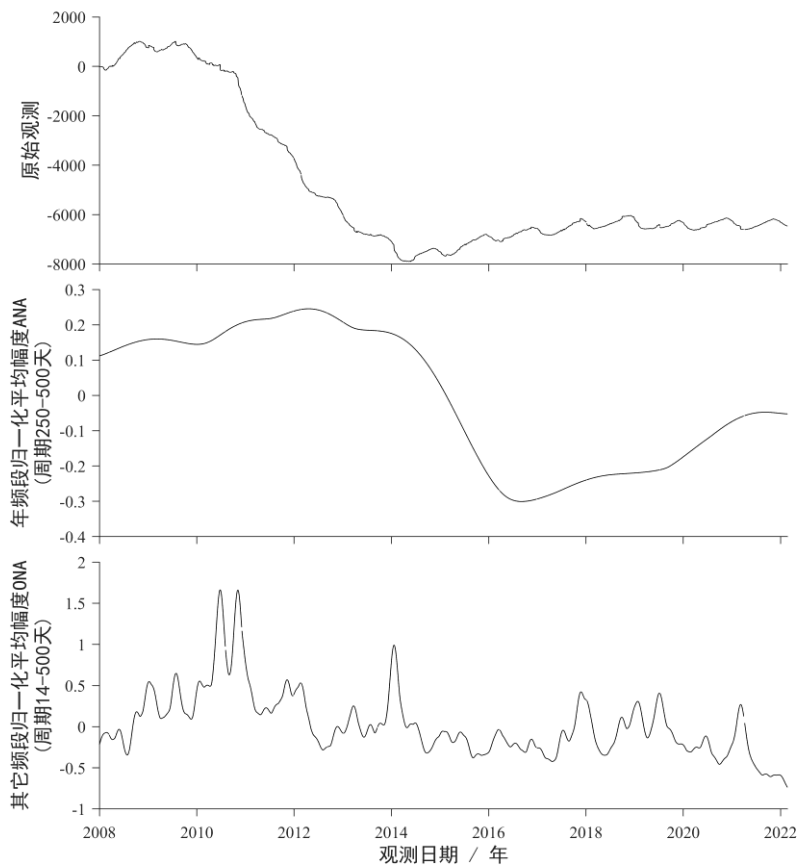


图 3 原始观测曲线及 ANA、ONA 信息曲线

软件默认参数划分 $M \geq 5$ 、6、7 三个震级档，由于观测期间台站周边无 7 级以上地震发生，因此该数据实际上仅在 $M \geq 5$ 和 $M \geq 6$ 两个震级档下针对 ANA 和 ONA 信息指标开展预测指标确定和效能评估。考虑地震中期预测和短临预测的业务需求，预测时窗默认参数设置为（7，14，30，60，。。。1080 天），最短为周尺度，最长接近 3 年，主要以月尺度为参数增量。预测空间范围默认参数设置为以台站为圆心的圆形区域，半径参数增量设为 50km，最大扫描半径为 500km。对于给定空间范围内地震次数少于 2 次的情况，出于结果稳定性和可信度的考虑不做分析。异常阈值的设置自动基于信息指标的变化范围，默认为 100 个等间隔参数。

基于默认的参数设置，程序会输出 ONA 信息对应的 5 级、6 级以上地震的最优时空参数、阈值等预测指标以及效能评估的扫描结果（图 4、图 5）。

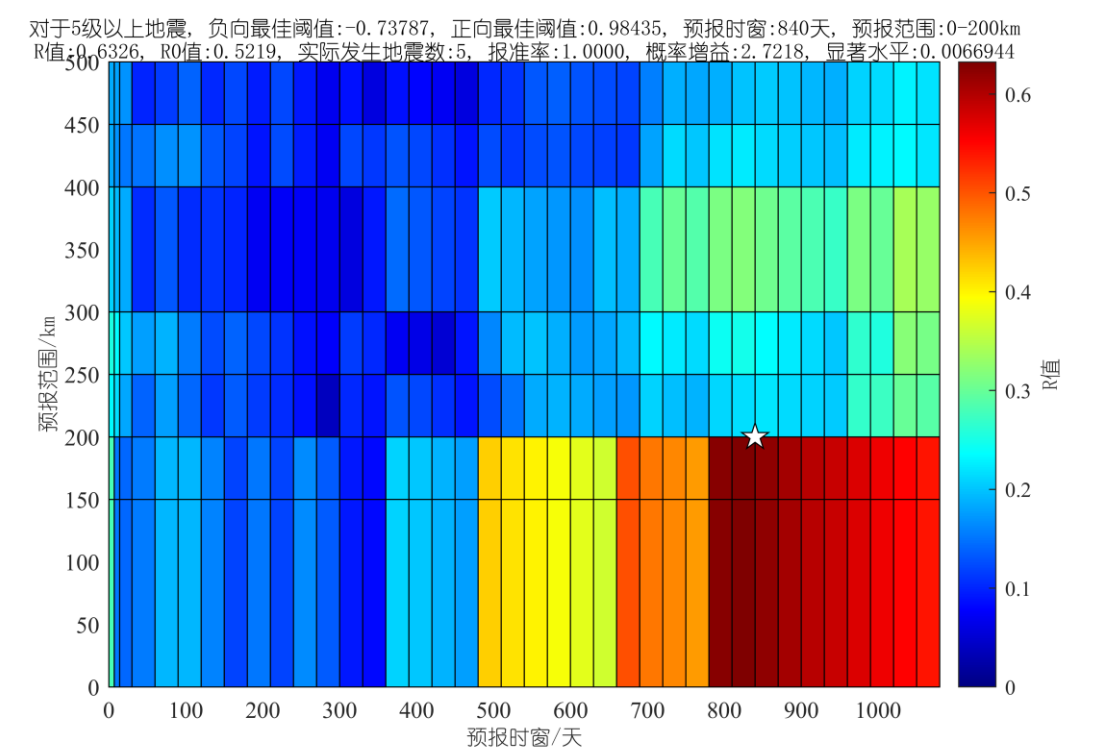


图 4 ONA 信息 5 级以上地震最优时空参数扫描结果

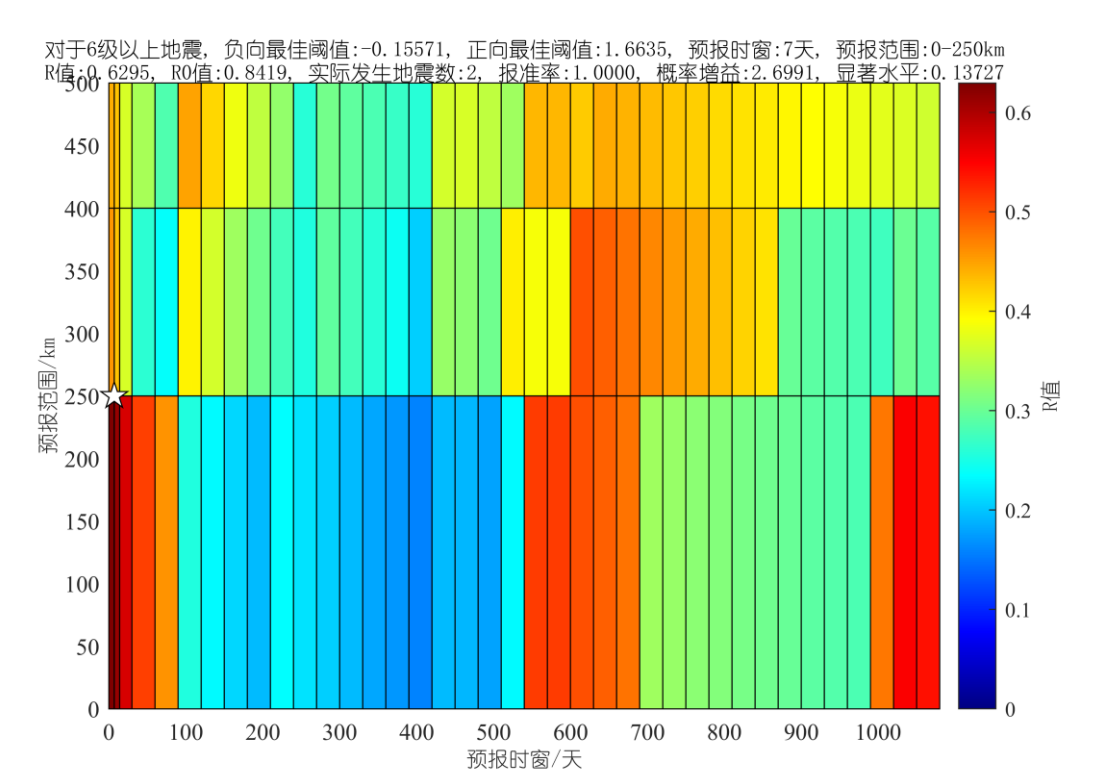


图 5 ONA 信息 6 级以上地震最优时空参数扫描结果

程序同样会输出 ONA 信息对应的 5 级、6 级以上地震的最优指标对应的 Molchan 图表（图 6、图 7）以及地震对应图（图 8、图 9）。

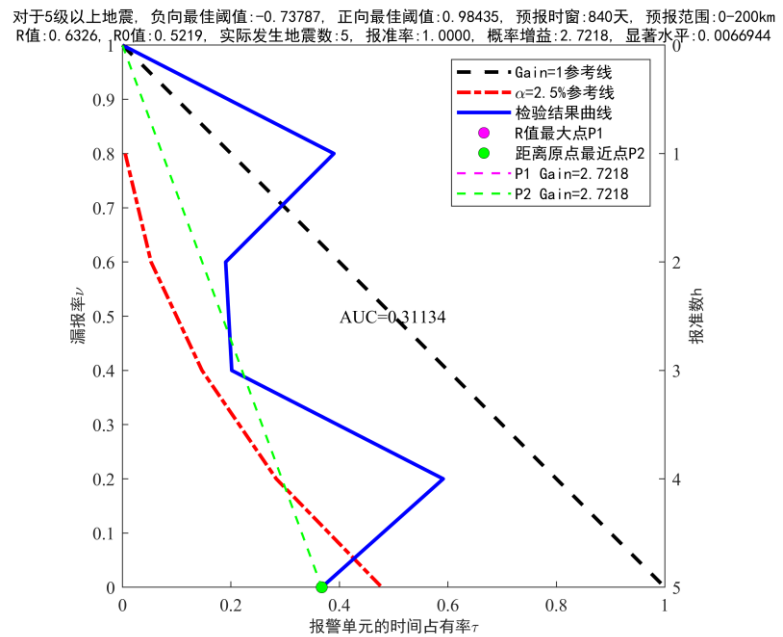


图 6 ONA 信息 5 级以上地震最优指标对应 Molchan 图表

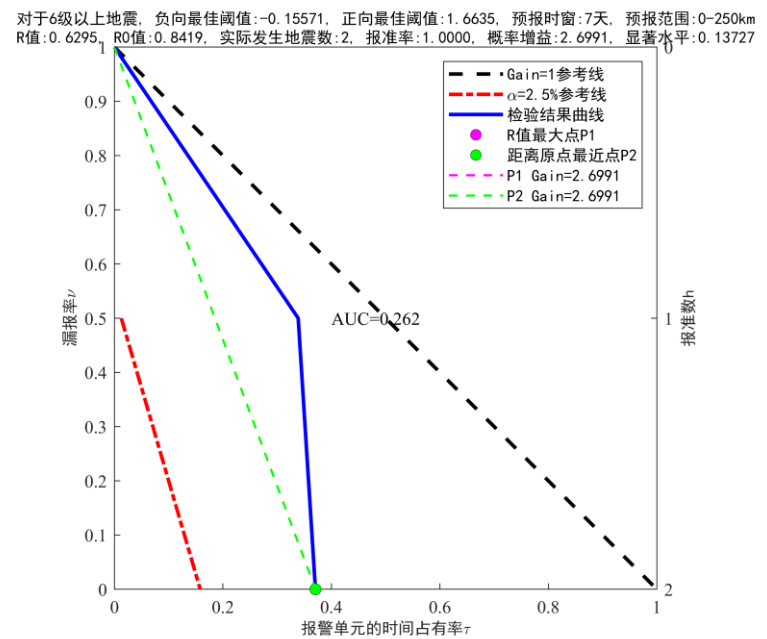


图 7 ONA 信息 6 级以上地震最优指标对应 Molchan 图表

对于5级以上地震，负向最佳阈值:-0.73787，正向最佳阈值:0.98435，预报时窗:840天，预报范围:0-200km
R值:0.6326，R0值:0.5219，实际发生地震数:5，报准率:1.0000，概率增益:2.7218，显著水平:0.0066944

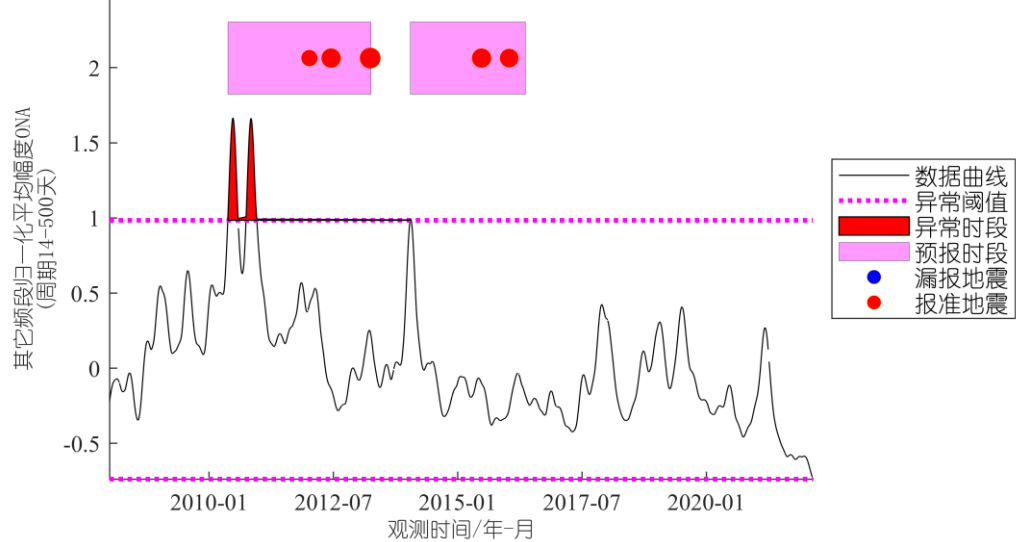


图 8 ONA 信息 5 级以上地震对应图

对于6级以上地震，负向最佳阈值:-0.15571，正向最佳阈值:1.6635，预报时窗:7天，预报范围:0-250km
R值:0.6295，R0值:0.8419，实际发生地震数:2，报准率:1.0000，概率增益:2.6991，显著水平:0.13727

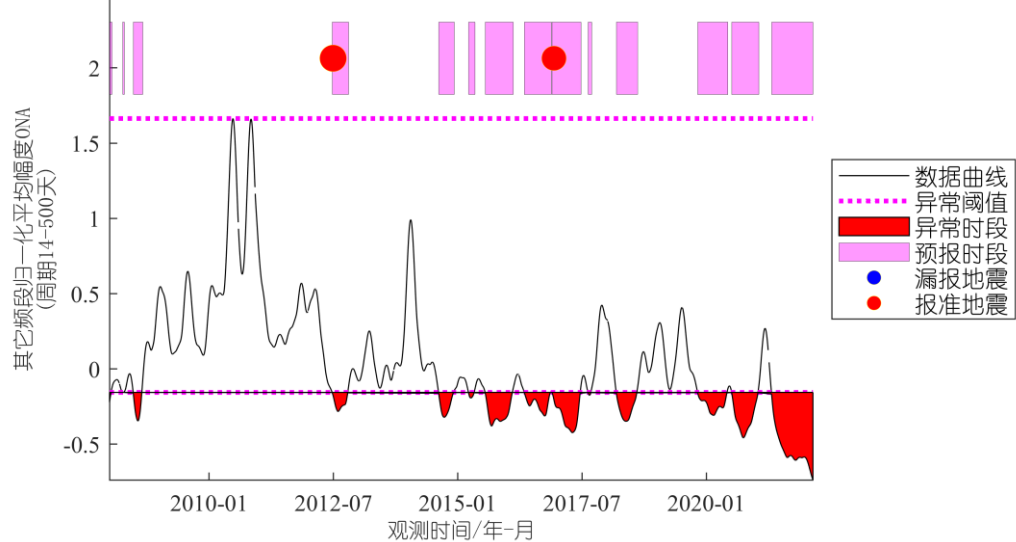


图 9 ONA 信息 6 级以上地震对应图

类似的，程序会输出 ANA 信息对应的 5 级、6 级以上地震的最优时空参数、阈值等预测指标以及效能评估的扫描结果（图 10、图 11），最优指标对应的 Molchan 图表（图 12、图 13）以及地震对应图（图 14、图 15）。

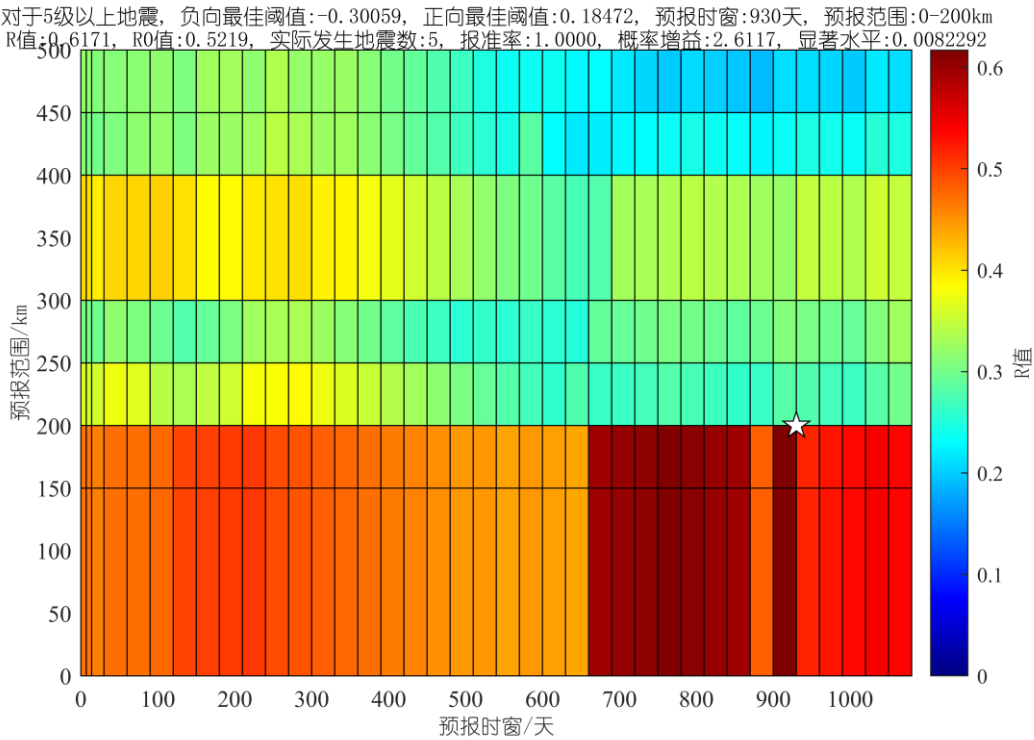


图 10 ANA 信息 5 级以上地震最优时空参数扫描结果

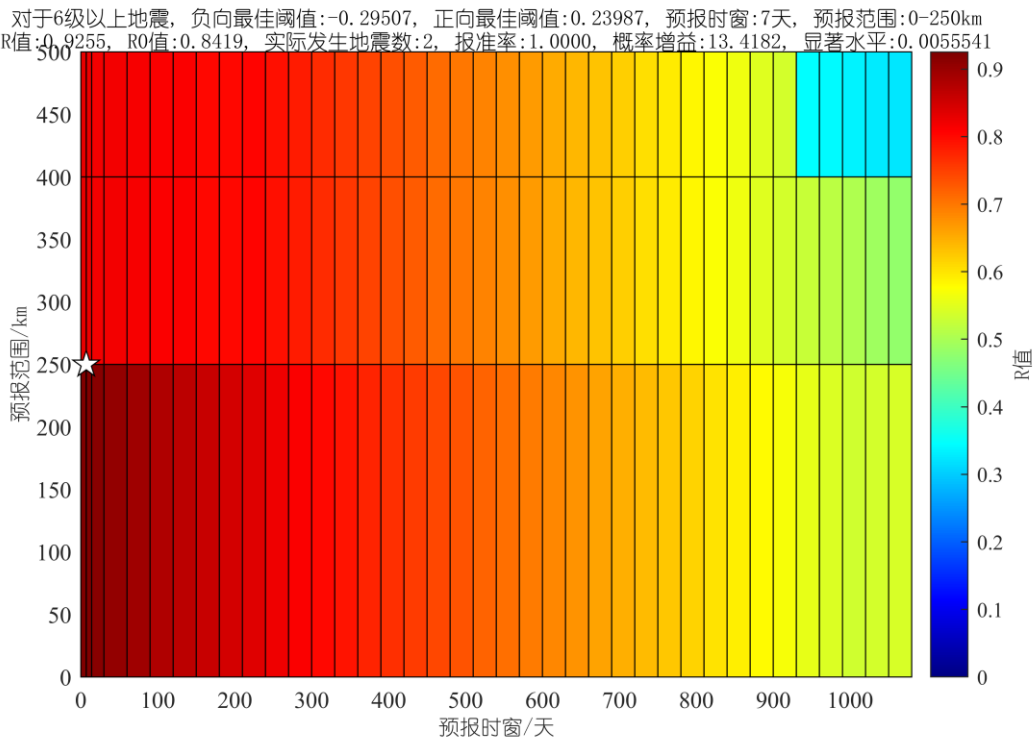


图 11 ANA 信息 6 级以上地震最优时空参数扫描结果

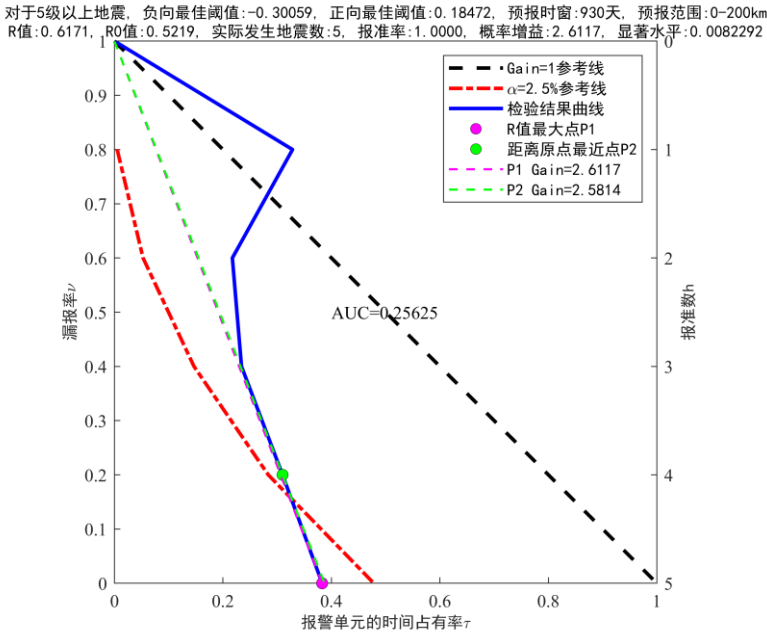


图 12 ANA 信息 5 级以上地震最优指标对应 Molchan 图表

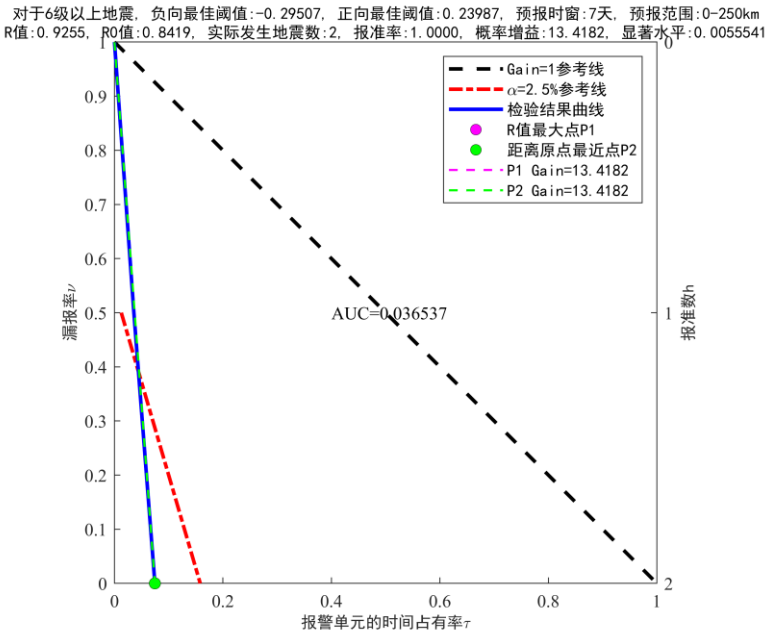


图 13 ANA 信息 6 级以上地震最优指标对应 Molchan 图表

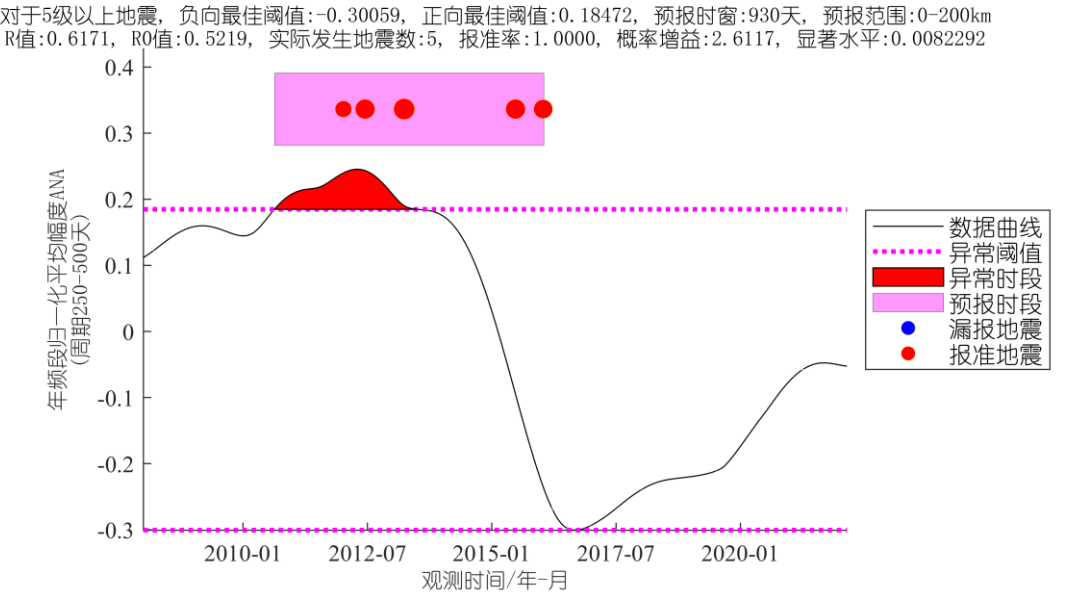


图 14 ANA 信息 5 级以上地震对应图

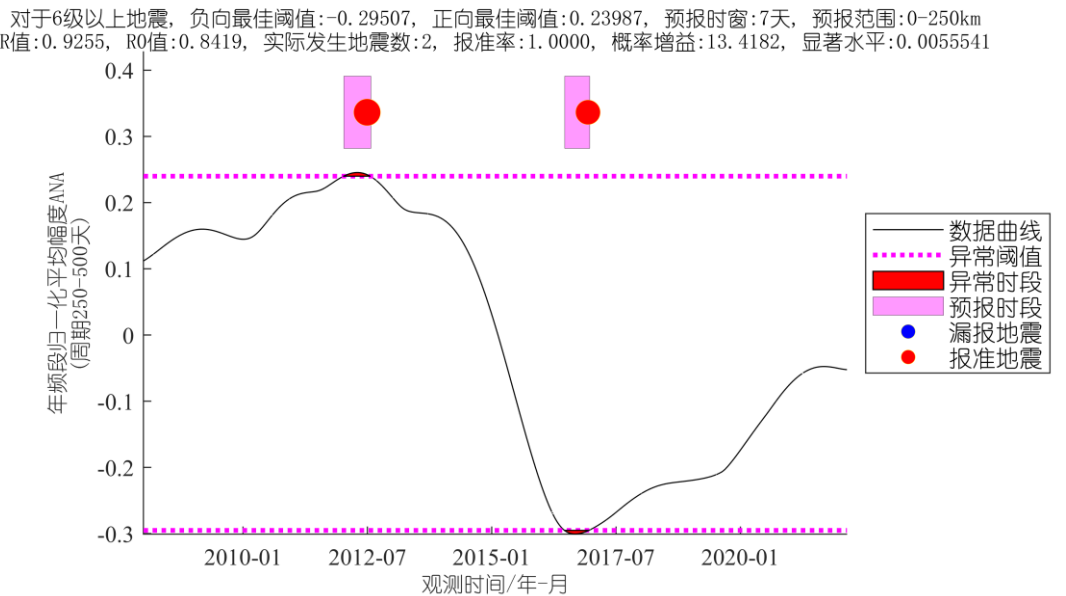


图 15 ANA 信息 6 级以上地震对应图

从对比结果来看，ANA 信息和 ONA 信息对于该台站周边 5 级以上地震最优的预报范围均为 200km 以内，基于最优预测指标对于观测期间发生的 5 次 5 级以上地震均可正确预测（5 次地震分别为 2012 年 1 月 8 日和硕 5.0 级、2012 年 6 月 15 日轮台 5.4 级、2013 年 3 月 29 日昌吉 5.6 级、2015 年 6 月 25 日托克逊 5.4 级和 2016 年 1 月 14 日轮台 5.3 级地震）。两种信息对应的最优指标其 R 值均大于 R0 值，统计意义上可信度较高，相比而言，ONA 信息最优预测指标对应的

预报占时率、显著性水平更低， R 值和概率增益更高，因此预测效能相对更优。

从 6 级以上地震的对比结果来看，ANA 信息和 ONA 信息最优的预报范围均为台站周边 250km 以内，基于最优预测指标对于观测期间发生的 2 次 6 级以上地震均可正确预测（2 次地震分别为 2012 年 6 月 30 日新源 6.6 级和 2016 年 12 月 8 日呼图壁 6.2 级地震）。由于地震次数较少，因此相应的 R_0 值较高，只有 ANA 信息对应的最优指标其 R 值大于 R_0 值，统计意义上可信度较高，ONA 信息最优指标对应的预报占时率较高，导致 R 值偏低，未通过显著性水平 2.5% 的检验。

从以上结果来看，对于新疆库尔勒台水平摆倾斜北南测项，其 ANA 和 ONA 信息指标的地震预测效能存在差异性，ONA 信息指标对于 5 级以上地震的预测效果相对更优，ANA 信息指标则在 6 级以上地震的预测中效果最佳。

四、其它说明

需要注意的是，由于震源特征、周边构造、台站环境、仪器性能等的差异，不同观测数据可能在地震前表现出不同的异常特征，需要不同类型的异常提取方法，不能因为某类方法效能不佳而直接认为该观测数据在地震预测中无法发挥作用。另外，对于基于概率统计获得的预测指标在使用过程中需要慎重，因为大部分指标都会面临地震样本数较少导致的稳定性问题，而且地震实况经常出现与概率预测不一致的情况，所以在震情研判过程中相关指标更适合作为参考，对于未来地震的预测还需综合多方因素。

后续将考虑基于数据库的干扰信息自动读入及影响剔除，进一步提升异常信度和效能。并将进一步针对空间筛选规则进行研究，发展自动化的构造相关筛选策略。