基于 FME 的跨带线性工程 DEM 接边分析

崔欣

对于大型线性工程的影像数据 ,跨带工程的数据接边是否满足要求是数据质检过程中的重中之重。传统检查方式是在接边处取采样点，收集采样点在两个投影坐标系下的高程信息，进行高差分析，以评估整个接边情况。但是这种利用样本描述整体的分析方式对样本高度依赖，受采样方式、采样密度等的影响极大，同时花费的时间和成本因人和手段而异，评估结果数据不稳定，容易遭到质疑。

FME 是高效数据批处理软件的代表，支持访问多种栅格数据格式，并含有丰富的栅格数据处理转换器。因此利用 FME 的高效数据处理性能对接边整体进行分析，分析标准化，避免了采样数量与影像重叠范围不匹配、采样方法、采样均匀度等主客观因素的影响，同时其结论更稳定。

制作处理模板

采用 FME Desktop 对跨带 DEM 数据接边分析需要进行的处理包括数据的坐标一致化、提取重叠区域数据、格网化、同名格网匹配、计算高差、摄取异常值、高差统计分析等过程，具体如下：

图 1 数据检查流程

第一步 坐标转换

首先进行坐标转换的原因是待检查的两块影像位于相邻的两个投影带中，采用两个不同的投影坐标系。为了进行重叠区域的分析，必须将两块影像坐标转换到同一个坐标系中。为了减少坐标转换带来的数据投影变形，将两块影像从各自的投影坐标系转换到 CGCS2000 地理坐标系。此步采用 Reprojector 转换器即

可满足需求。

第二步 提取影像边界

为了计算影像的接边区域，即两块影像的重叠区域，可从两块影像有效数据的边界多边形重叠范围计算得到。对多边形重叠区域计算的效率要比直接逐像素判断两块影像的重叠范围更高效，因此需要先提取两块影像各自的边界多边形。

由于线性工程的影像多为窄长条状，数据并不是布满整个最小外接矩形，因此需要提取的是有效数据的边界多边形，排除无效数据的干扰，同时也可以提升数据处理分析的效率。

对于支持 Nodata 的影像格式同时有效利用 Nodata 字段的数据，可直接利用 FME 的 RasterExtentsCoercer 转换器提取边界，边界类型为数据边界。

对于没有使用 Nodata 字段赋值无效数据或不支持的 Nodata 的影像数据，需要了解其对无效数据的赋值数值，利用 RasterBandNodataSetter 转换器让模板识别出无效数据。注意为了便于通过该转换器的数据可以利用 RasterExtentsCoercer 转换器提取边界多边形，应将其参数“代替单元值”设置为“是”。

第三步 计算重叠区域

有效数据的边界多边形已经取得，通过 AreaOnAreaOverlayer 转换器可以对两个多边形的重叠情况进行判断，记录在重叠数量字段中，再通过 TestFilter 转换器提取重叠数量大于等于 2 的区域，即得到两块影像的重叠区域范围。

但是由于 RasterExtentsCoercer 转换器获得的边界多边形是沿着影像栅格

边缘精确绘制的，因此该多边形通常含有数量巨大的节点，会对计算机内存已经计算效率有一定的影响。所以需要对节点数量进行优化，可以使用 Generalizer 转换器实现此目的。同时由于通常在影像数据的边缘处，数据通常质量欠佳，因此在进行接边分析时，可以基于影像边界适当缩小，取得数据分析的范围。利用 GeographicBufferer 转换器可以基于重叠区域边界进行向内缓冲一个半格网尺寸，获得接边分析的数据范围。

第四步 裁剪取得接边区影像

根据已经获得的重叠区域多边形，分别对两块原始影像进行裁剪，只保留需要计算分析的数据区域，为后续逐像素的计算分析做准备。利用 Clipper 转换器即可实现该需求。

第五步 影像格网化

由于 DEM 中的高程信息都存储在每个格网的 band 信息中，默认在模板数据流中无法直接访问每个单元的信息，因此需要对影像数据进行格网化。利用 RasterCellCoercer 转换器，定义输出单元几何结构为点并提取波段值到属性中，可以实现影像格网化需求，以访问每个栅格单元的高程信息。

由于 RasterCellCoercer 将高程信息自动提取到 band 列表中，因此需要将高程显示提取到固定的属性字段中。新建一个高程字段复制存储相关信息即可。

第六步 同名格网匹配

以一个数据集为依据，对该数据集中的每个格网，寻找其在另一个数据集中的最邻近点，将这对格网当作同名格网，即同一位置在两块影像上的格网单元。

利用 NeighborFinder 转换器可在两个数据集中寻找最邻近点。由于同名格网的邻近距离不会大于 1 个格网尺寸，因此可以设置为约束条件，即最大距离参数不超过一个格网尺寸。

为了便于后期计算两个同名格网的高程差值，需要对两格网的属性进行合并，并对候选集的数据进行前缀标记，以防止属性冲突时信息的丢失风险。

第七步 计算高差

新建高差字段存储同名格网高程的差值。

由于在影像格网化及同名格网匹配的过程中，可能会有些数据略微超出了接边分析的范围，因此需要追加一次空间过滤，将边缘可能有问题的数据提前过滤掉。

第八步 分离异常值

当同名格网的高差大于一个格网尺寸时，则认为该点存在接边异常的可能性，因此，需要将此类点分流出来，并单独输出到一个矢量文件，便于后续的确认分析。

第九步 接边统计分析

对认为没有接边异常的格网进行统计分析，获取这些格网单元的高差最大值、高差最小值、高差均值、高差标准差、高差众数等，以及参与统计分析的格网数量。

为了进行对比，同时对接边范围内包括上一步获取的潜在异常值的所有单元进行同样的统计分析，获取接边统计分析情况。这是因为，当接边高差是整体存

在偏移，且偏移大于一个格网单元时，将可能没有高差小于一个格网的单元，进而也就无法获知整个接边区的统计信息。为了对这种存在整体偏移的接边情况兼容，有必要对两组分离的数据进行全部统计，这样统计结果可以明显指出整体偏移量。

应用与分析

样例数据：某横跨高斯三度投影带 30-35 的长输管道工程沿线的 DEM 数据，有效数据的长度约 1700 公里、宽度 5 公里，共计将近 9000 平方公里。原始数据按照不同的投影带分成 6 块，接边处有重叠区域。在该数据使用之前，需要对数据的接边情况进行检查，便于后续整个工程对于数据使用。

整个接边分析的模板如下图所示：

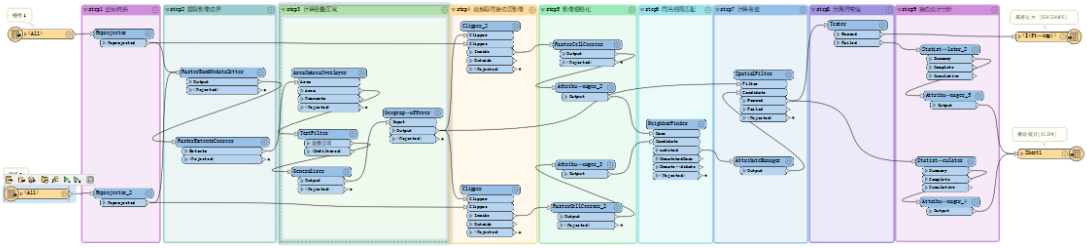


图 2 数据分析模板

下图所示其中两个带的影像数据。红色框的范围内是一个带的数据，其余部分是另一个带的数据。

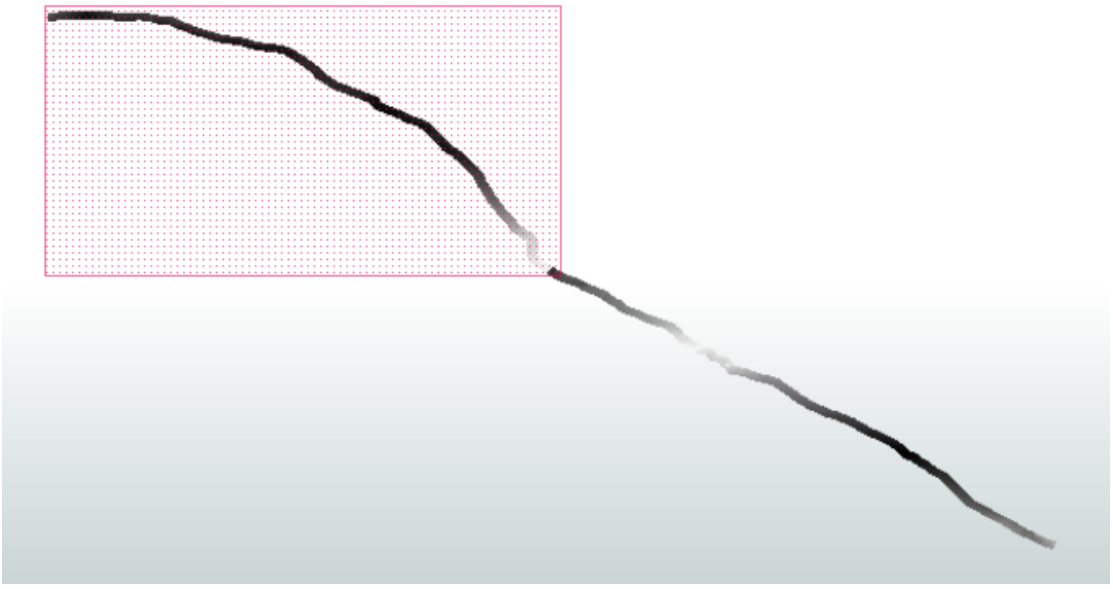
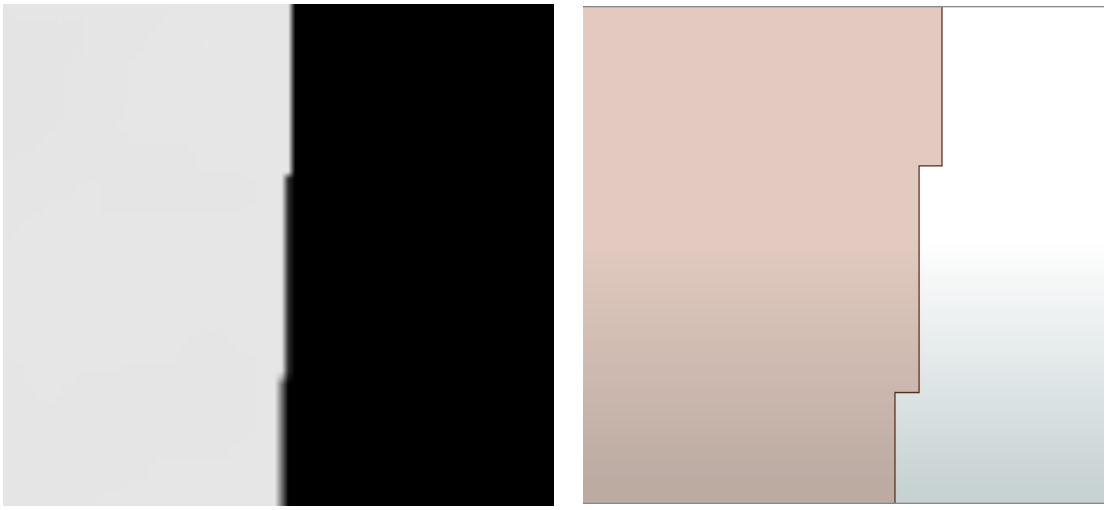


图 3 数据状态

RasterExtentsCoercer 转换器提取的边界多边形与影像的真实情况对比如下图所示，左侧（a）是原始影像，右侧（b）是获取的边界多边形。



(a) (b)

图 4 影像边界

边界多边形叠置分析得到的重叠区域如下图所示。

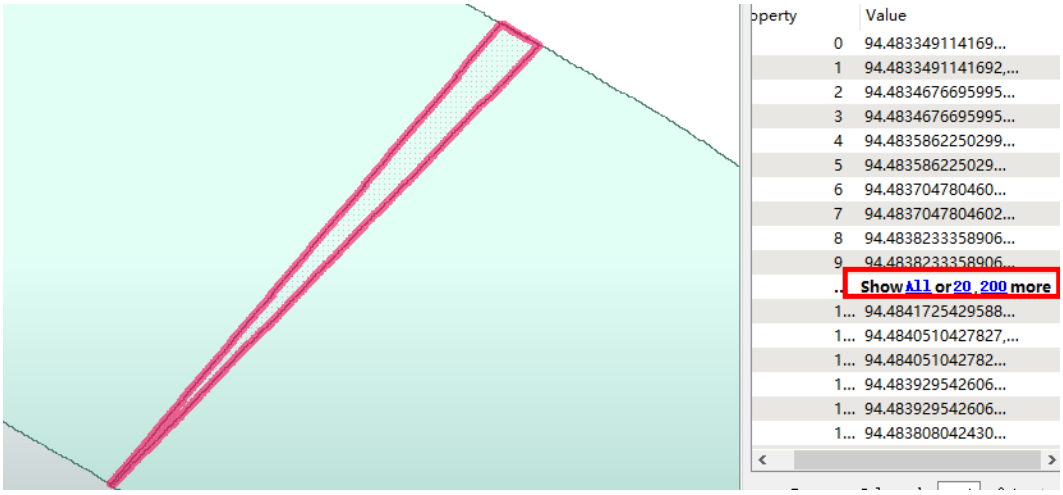


图 5 叠置分析结果

缩减多边形顶点，并进行向内缓冲后的重叠区域多边形如下图所示。

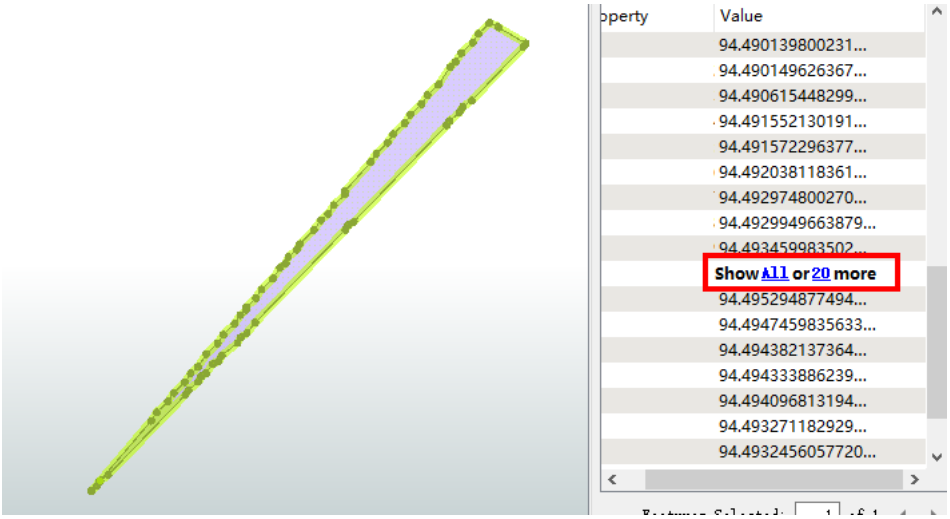


图 6 分析用的重叠区域

对其中两个接边统计结果如下图所示。30-31 接边由于只输出了全部统计分析结果，观察其标准差小于 1 而均值为 34.9，说明整体存在偏移，数据未通过验收，需要整改。31-32 接边的两条统计记录说明未存在高差超出限值的单元，其标准差 0.335 说明数据离散度不高，结合其他统计数据可以得出该处 DEM 数据的接边可以满足后续生产上的需要。

接边处	min	max	median	count	mean	stdev	mode
30-31_Total	31	39	35	233111	34.948	0.468	35
31-32	-1	1	0	7935	-0.02	0.335	0
31-32 Total	-1	1	0	7935	-0.02	0.335	0

图 7 接边统计结果

下图为 FME2019beta 版计算上述统计结果花费的时间，其效率相对人工统计计算的花费提升是巨大的，且该统计分析的模板具有很强的通用性。

```
Translation was SUCCESSFUL with 12 warning(s) (233112 feature(s) output)
FME Session Duration: 7 minutes 43.6 seconds. (CPU: 397.6s user, 32.6s system)
```

```
Translation was SUCCESSFUL with 11 warning(s) (2 feature(s) output)
FME Session Duration: 6 minutes 20.8 seconds. (CPU: 331.1s user, 16.7s system)
```

图 8 FME 数据分析耗时

展望

DEM 接边处的高差分析说明统计分析可以确定两块跨带 DEM 接边重叠区域的值是否可以满足数据合并使用的需求，同理 DOM 的接边检查应该在波段信息上也具有一定相关性，未来可以考虑利用 DOM 的波段信息，进行适当的相关性统计分析，以确定接边情况是否满足数据验收要求。