

（一）ArcGIS 和 Pro 配置环境

【说明】ArcGIS Pro 安装依赖库比 ArcGIS 简单

ArcGIS Pro: python3 ArcGIS Pro2.5 对应 python3.6

ArcGIS: python2.7

一、版本需求

本工具箱要求 **ArcGIS 版本不低于 10.1**，ArcGIS Pro 无版本要求。

二、ArcGIS Pro 安装其他 Python 库

开始菜单——ArcGIS——Python Command Prompt

类似于 Anaconda 安装 python 库一样，如果速度很慢可切换到国内源：

```
pip install scikit-learn -i https://pypi.tuna.tsinghua.edu.cn/simple --user
```

```
pip install seaborn -i https://pypi.tuna.tsinghua.edu.cn/simple --user
```

三、ArcGIS 安装其他 Python 库

1、首先检查安装 ArcGIS 时，Python2.7 安装路径是否修改，如修改则应重新安装，Python2.7 安装路径保持默认。

检查是否默认：在 “C:\python27” 路径是否存在，若存在，则未修改路径，可以进行下面的步骤。

2、打开 cmd，进入 C:\Python27\ArcGIS10.X\Scripts 路径下，使用

```
pip install scikit-learn -i https://pypi.tuna.tsinghua.edu.cn/simple
```

```
pip install seaborn -i https://pypi.tuna.tsinghua.edu.cn/simple
```

（二）SVM-LSM 工具箱的使用

一、工具箱安装

在 ArcGIS——“目录”界面——“文件夹连接”——添加本工具箱所在路径（图 1）即可。



图 1 添加工具箱

二、工具箱介绍

本工具箱旨在提供一个基于面数据的支持向量机的滑坡易发性制图工具箱，工具箱包括“1 影响因子制作”、“2 数据集制作与因子筛选”、“3 模型训练及预测”三个子工具箱和 12 个工具，如图 2 所示。本工具箱的开发基于 *Arcpy* 与模型构建器实现。需要在 ArcGIS 或 ArcGIS Pro Python 环境下安装 *scikit-learn* 与 *seaborn* 依赖库。下文工具箱具体使用方法以配套案例数据为例。



图 2 SVM-LSM 工具箱示意图

1、影响因子制作

(1) dem 相关因子计算

【功能】

根据研究区 DEM 数据自动计算其他地形因素，如坡度、坡向、曲率、平面曲率、剖面曲率、地形起伏度、地表粗糙度、地形湿度指数（TWI）等。注意：DEM 数据必须为 UTM 投影坐标系。计算平面曲率时必须计算坡向；计算剖面曲率、地表粗糙度或地形湿度指数时必须计算坡度。

【输入】

- ① DEM 数据（UTM 坐标系）：必填。研究区 DEM 栅格数据（UTM 坐标系下）。
- ② 工作空间：必填。输出文件存放目录，该工具会自动在此路径下生成一个“*dempro*”的文件夹存放数据。
- ③ 计算因子选择：可选。根据需要，自主选择需要计算的因子并选中，填写因子名；如不需要计算，则不选择，因子名不填。因子名为该因子输出后保存名称。注意：计算平面曲率时必须计算坡向；计算剖面曲率、地表粗糙度或地形湿度指数时必须计算坡度。

【输出】

由 DEM 计算得到的坡度、曲率、平面曲率、剖面曲率、地形起伏度、地表粗糙度、地形湿度指数（TWI）栅格数据。

为方便后续批量裁剪，建议将 *dempro* 文件夹内数据移动至上一级文件夹下。



图 3 dem 相关因子计算

(2) 线数据转连续型因子

【功能】

将研究区线矢量数据自动转换为连续型栅格数据，如道路转换为距道路的距离、断层转换为距断层的距离、水系转换为距水系的距离等。其中，生成的栅格数据分辨率为 30m，转换原理采用欧氏距离。注意：线矢量数据必须为 UTM 投影坐标系。

【输入】

① 线矢量数据：必填。所计算的线矢量数据，如：

道路_UTM：研究区道路数据（UTM 坐标系下）。如不计算此项，则不添加即可。

断层_UTM：研究区断层数据（UTM 坐标系下）。如不计算此项，则不添加即可。

水系_UTM：研究区水系数据（UTM 坐标系下）。如不计算此项，则不添加即可。

② 文件保存路径：必填。输出文件存放目录，生成后直接存放在该文件夹下。

【输出】

距道路的距离、距断层的距离、距水系的距离 栅格数据。

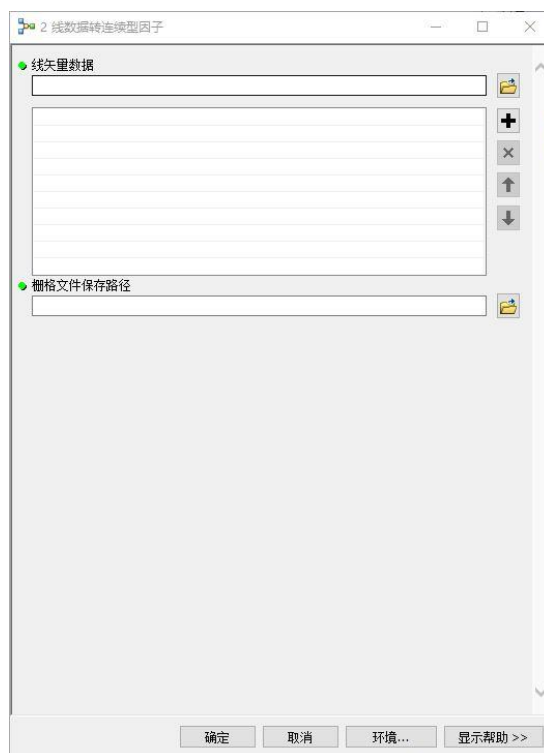


图 4 线数据转连续型因子

(3) 降雨量数据处理

【功能】

本工具针对从 NASA (<https://gpm.nasa.gov/>) 下载的逐月或逐天降雨量数据（.nc4 格式）数据转换为 30m 的 tif 栅格数据。转换后仍为逐月或逐天的.tif 栅格数据。

【输入】

- ① 降雨量数据（.nc4）所在文件夹路径：必填。从 NASA 下载的逐月或逐日降雨量数据所在的文件夹。
- ② 输出坐标系：必填。转换后的.tif 数据坐标系，建议选择 UTM 投影坐标系。

【输出】

与.nc4 文件对应的降雨量.tif 文件。最后在输入路径下新建一个“UTM_30m”的文件夹存放生成的栅格数据。

后续使用时应使用栅格计算器根据需要进行叠加，本案例中后续使用年降雨量，因此使用栅格计算器将所有.tif 累加（严格使用栅格计算器操作，切勿键盘输入“+”等），得到年降雨量。

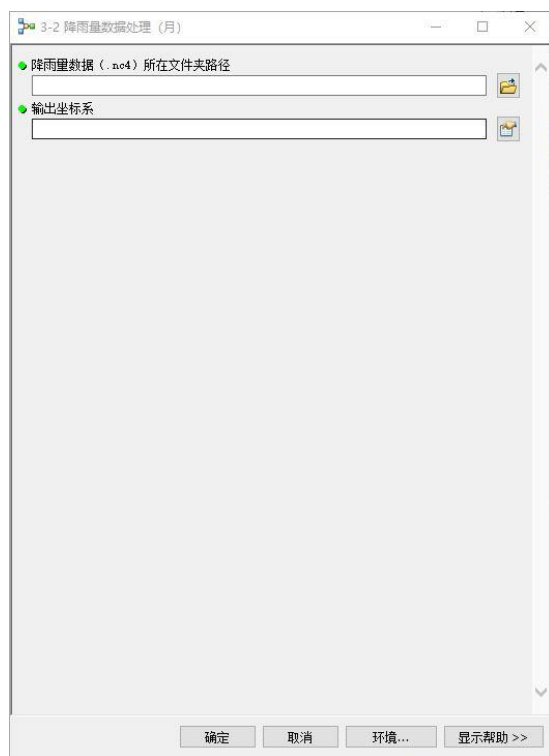


图 5 降雨量数据处理（月）

（4）批量裁剪各因子层

【功能】

本工具用于根据研究区矢量数据批量裁剪各因子层栅格数据，得到研究区因子层数据。本工具仅需给定栅格因子所在文件夹，自动迭代选择`.tif` 文件进行裁剪。注意：矢量数据和栅格数据必须是 UTM 投影坐标系，栅格数据分辨率保持一致。

【输入】

- ① 研究区矢量数据（UTM 坐标系）：必填。选择研究区矢量数据（`.shp`）文件，注意该矢量数据已投影到 UTM 投影坐标系下。
- ② 输入栅格影像所在文件夹（栅格文件 UTM 坐标系）：必填。需要裁剪的栅格数据所在文件夹，会自动识别`.tif` 文件并迭代裁剪，注意所有栅格数据已投影到 UTM 投影坐标系下。
- ③ 使用输入要素裁剪几何：可选。**不建议选！**选中 - 使用选定要素类的几何裁剪数据。可以增加输出的像素深度，因此需要确保输出格式可以支持适当的像素深度。未选中 - 使用最小外接矩形裁剪数据。
- ④ 保持裁剪范围：可选。**强烈建议，一定要选！**选中 - 调整列数和行数并对像

素进行重采样，以便完全匹配指定的裁剪范围。未选中 - 维护与输入栅格相同的像元对齐方式并相应调整输出范围。

⑤ 输出数据文件夹：必填。裁剪后的研究区因子层栅格数据存放文件夹，生成后直接存放在该文件夹下。

【输出】

裁剪后的研究区因子层栅格数据。



图 6 批量裁剪各因子层

2、数据集制作与因子筛选

(1) 非滑坡点数据生成

【功能】

本工具用于在研究区矢量数据图层范围内生成非滑坡点数据。原则：在给定滑坡样本点一定缓冲区范围外随机选取相同数量的非滑坡样本点。注意：研究区矢量数据和滑坡点矢量数据必须是 UTM 投影坐标系，得到的非滑坡点矢量数据默认与滑坡点矢量数据坐标系保持一致。

【输入】

① 滑坡点要素 (UTM 坐标系)：必填。选择研究区滑坡点矢量数据 (*.shp*) 文件，

注意该矢量数据已投影到 UTM 投影坐标系下。

② 距滑坡点距离：必填。在滑坡点多少米缓冲区范围外选取非滑坡样本点。

③ 研究区矢量文件（UTM 坐标系）：必填。选择研究区矢量数据（.shp）文件，注意该矢量数据已投影到 UTM 投影坐标系下。

④ 点数：必填。滑坡点个数，也即最后生成的非滑坡点个数。

⑤ 输出位置：必填。非滑坡点矢量数据文件输出文件夹路径。非滑坡点矢量数据文件输出文件名默认为 *non_landslide_point.shp*，注意该矢量数据已投影到 UTM 投影坐标系下。

⑥ 输出坐标系：可选。非滑坡点矢量数据文件输出坐标系，默认与滑坡点矢量数据坐标系保持一致。

【输出】

研究区非滑坡样本点矢量数据。



图 7 非滑坡点数据生成

(2) 数据样本制作

【功能】

此工具根据矢量点数据生成多通道块样本栅格数据。原则：利用矢量点数据（.shp）制作缓冲区并逐个要素裁剪多通道栅格数据（.tif），得到各要素的单个裁

剪结果并以“*FID*”值命名。注意：矢量点数据（*.shp*）和多通道栅格数据（*.tif*）均在同一 UTM 投影坐标系下。得到的多通道块样本栅格数据默认与多通道栅格数据（*.tif*）坐标系保持一致。

【输入】

- ① 输入矢量点要素：必填。选择研究区滑坡点矢量数据文件（*.shp*）或非滑坡点矢量数据文件（*.shp*），注意该矢量数据已投影到 UTM 投影坐标系下。
- ② 缓冲区距离：必填。以滑坡点多少米缓冲区制作滑坡样本点。
- ③ 多通道因子层数据：必填。所使用的各因子层合并后的多通道栅格数据。
- ④ 裁剪后数据存放目录：必填。滑坡与非滑坡样本数据集存放路径。
- ⑤ 样本标签类别（landslide or non-landslide）：必填。数据集存放文件夹名称。
- ⑥ 使用输入要素裁剪几何：可选。**不建议选！**选中 - 使用选定要素类的几何裁剪数据。可以增加输出的像素深度，因此需要确保输出格式可以支持适当的像素深度。未选中 - 使用最小外接矩形裁剪数据。
- ⑦ 保持裁剪范围：可选。**强烈建议，一定要选！**选中 - 调整列数和行数并对像素进行重采样，以便完全匹配指定的裁剪范围。未选中 - 维护与输入栅格相同的像元对齐方式并相应调整输出范围。

【输出】

研究区滑坡与非滑坡块数据样本。



图 8 数据样本制作

(3) 数据集划分

【功能】

此工具根据滑坡点数量及上一步生成的数据样本，按照测试集比例划分训练集和测试集，并以 *txt* 保存划分结果。

【输入】

- ① 样本所在目录：必填。选择上一步滑坡与非滑坡样本存放路径。
- ② 滑坡点个数：必填。滑坡点个数。
- ③ 测试集比例（如 0.3）：必填。测试集占总样本的比例，1 减该值为训练集占总样本的比例。

【输出】

生成三个文本文件，分别为所有样本路径及标签 (*XXX.txt*)，所有训练样本路径及标签 (*XXXtrain.txt*)，所有测试样本路径及标签 (*XXXtest.txt*)。

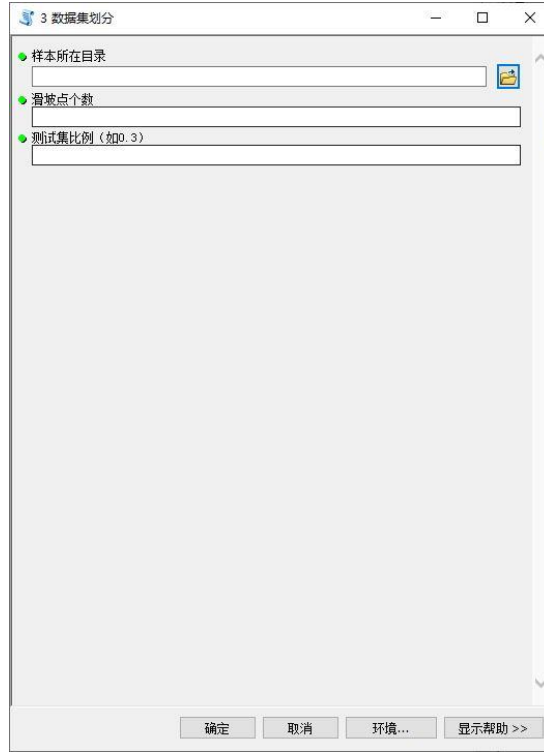


图 9 数据集划分

(4) PCC 与 IGR 计算

【功能】

此工具基于生成的数据样本与所有样本路径及标签文件，计算各影响因子层的皮尔逊相关系数与信息增益比。皮尔逊相关系数代表的因子层之间的相关性大小，其值位于 $[-1, 1]$ 之间，相关性较大的因子应考虑剔除。信息增益比代表各因子层对滑坡发生的贡献大小，其值大于 0 说明对滑坡发生有贡献，值越大，贡献越大。

皮尔逊相关系数：

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

式中， r 表示皮尔逊相关系数， X_i 和 Y_i 分别代表第 i 个因子层， \bar{X} 和 \bar{Y} 分别表示 X 与 Y 的平均值。

信息增益比：

$$IGR(D, A) = \frac{IG(D|A)}{Splitinfo_A(D)}$$

D 为训练集中某一滑坡相关因子值样本标签，即是否为滑坡， A 为训练集中某一滑坡相关因子值。 $IG(D|A)$ 表示某一滑坡相关因子值 A 对应的类别标签 D 的信息增益， $Splitinfo_A(D)$ 表示某一滑坡相关因子值 A 关于类别标签 D 的信息熵。即：

$$IG(D|A) = Infor(D) - Infor(D|A)$$

$$Splitinfor_A(D) = - \sum_{i=1}^m \frac{|D_i|}{|D|} \times \log_2 \left(\frac{|D_i|}{|D|} \right)$$

其中 $Infor(D|A)$ 为条件熵， $Infor(D)$ 为信息熵，计算公式如下：

$$Infor(D|A) = \sum_{j=1}^n \frac{|D_j|}{|D|} \times Infor(D_j)$$

$$Infor(D) = - \sum_{j=1}^n p(x_j) \times \log_2 p(x_j)$$

【输入】

- ① 数据集所在目录：必填。选择上一步滑坡与非滑坡样本存放路径。
- ② 结果保存路径：必填。PCC 与 IGR 结果存放路径。
- ③ 输入因子层顺序：必填。因子层栅格数据叠加顺序，需要与“待预测影像生成”中“因子层叠加顺序”保持一致。

【输出】

生成【*PCC.txt*, *PCC.png*, *PCC.svg*, *IGR.txt*, *IGR.png*, *IGR.svg*】六个文件。

【**.txt*】PCC 或 IGR 结果 txt 文件。

【**.png*】PCC 或 IGR 结果图 png 文件。

【**.svg*】PCC 或 IGR 结果图可编辑矢量文件。

本工具得到结果后，应综合考虑 PCC 与 IGR 结果，剔除冗余因子。

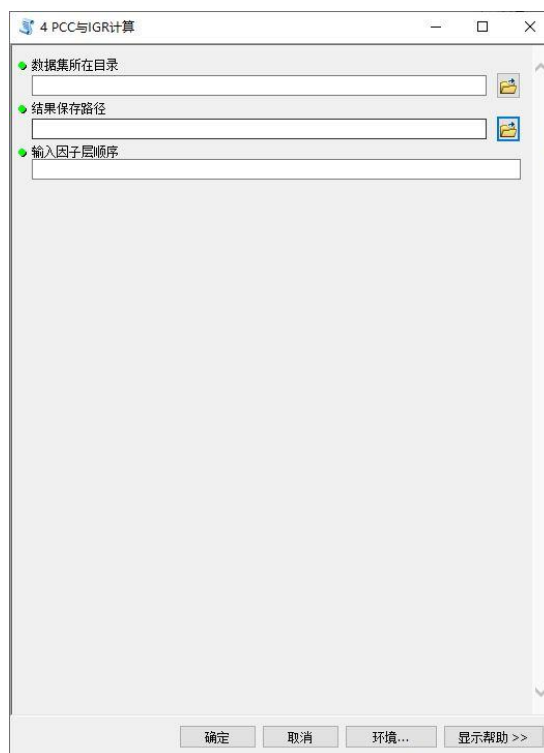


图 10 PCC 与 IGR 计算

3、模型训练及预测

(1) 待预测影像生成

【功能】

此工具根据各因子层栅格数据生成待预测影像多通道栅格数据。用于后续数据样本制作及易发性图预测。注意：所有因子栅格数据（.tif）均在同一 UTM 投影坐标系下。得到的多通道栅格数据默认与各因子栅格数据（.tif）坐标系保持一致。

【输入】

- ① 各因子层所在路径：必填。选择研究区各影响因子栅格数据所在目录，注意所有栅格数据已投影到 UTM 投影坐标系下。
- ② 因子层叠加顺序：必填。以什么样的顺序叠加因子层栅格数据，需要与上一步所选路径下各因子层名称保持一致。
- ③ 待预测影像保存文件：必填。叠加后的因子层多通道栅格数据所在文件夹。本工具会自动在所选文件夹下创建“Factors_所选因子层数量_mapping.tif”的栅格文件。

【输出】

研究区待预测影像多通道栅格数据，文件名为“Factors_所选因子层数量_mapping.tif”。

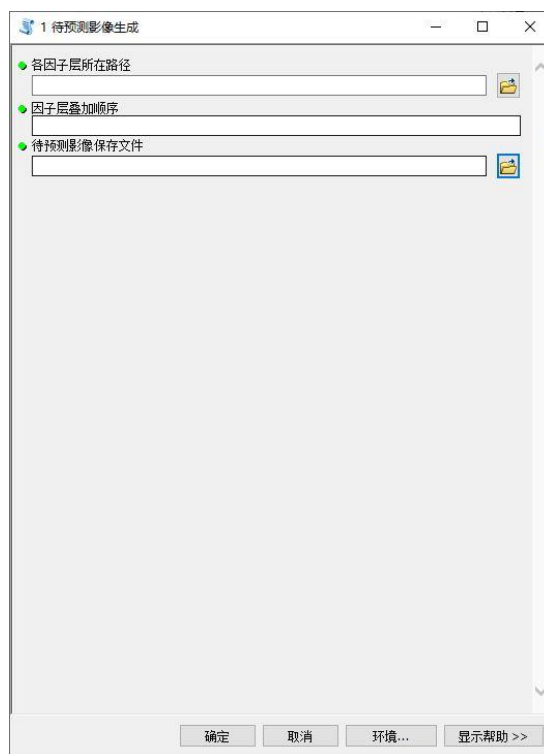


图 11 待预测影像生成

(2) SVM 模型训练及精度评定

【功能】

此工具用于生成各组参数下的 SVM 模型并给出模型精度的评定结果。SVM 核函数本工具默认为径向基核函数 (RBF)，需要调节的参数为 γ 与惩罚因子 C ，使用的调参方法为网格搜索算法。

【输入】

- ① 数据集样本所在目录：必填。选择生成的滑坡与非滑坡样本存放路径。
- ② 模型保存路径：必填。选择各组参数训练得到的 SVM 模型的存放路径，该路径下会以“g_g 值_C_C 值”新建不同参数文件夹存放结果。
- ③ γ 参数取值：必填。SVM 径向基核函数中 γ 参数取值。
- ④ C 参数取值：必填。SVM 径向基核函数中 C 参数取值。
- ⑤ 数据集行数：必填。数据集样本的行数。
- ⑥ 数据集列数：必填。数据集样本的列数。
- ⑦ 数据集通道数：必填。数据集样本的通道数。

【输出】

【parameter_result_txt.txt】存放各组参数下的 SVM 模型在测试集上的 AUC 值、测试集精度 Test_acc、训练集精度 Train_acc 及两者之差。可用于最优模型的选择。

【parameter_result_png.png 与 parameter_result_png.svg】txt 文件绘图展示。图中，圆形大小代表 AUC 值，圆形越大，AUC 值越高，模型精度越好；圆形颜色代表在训练集和测试集上的精度之差，超过 0.5 的以 0.5 计，差值越大，表明模型过拟合程度越高，泛化性能越差。

【\g_0.02_C_0.5】: *gamma* 为 0.02，*C* 为 0.5 的 SVM 结果存放路径。

【\g_0.02_C_0.5\SVM_g_0.02_C_0.5.model】*gamma* 为 0.02，*C* 为 0.5 时训练好的 SVM 模型。

【\g_0.02_C_0.5\SVM_train_result_txt.txt】*gamma* 为 0.02，*C* 为 0.5 时 SVM 模型在训练集上的预测结果。

【\g_0.02_C_0.5\SVM_test_result_txt.txt】*gamma* 为 0.02，*C* 为 0.5 时 SVM 模型在测试集上的预测结果。

【\g_0.02_C_0.5\evaluate_result.txt】*gamma* 为 0.02，*C* 为 0.5 时 SVM 模型在测试集上的各类评价指标，如混淆矩阵，准确度，精度，F1 值，AUC 值等。

【\g_0.02_C_0.5\ROC.png】*gamma* 为 0.02，*C* 为 0.5 时 SVM 模型在测试集上的 ROC 曲线及 AUC 值。

根据【parameter_result_txt.txt 与 parameter_result_png.png】选取 AUC 值较高且差值较小的模型作为最优模型用于易发性图预测。



图 12 SVM 模型训练及精度评定

(3) 易发性图预测

该功能包括两个工具，分别为“易发性图预测（单进程）”和“易发性图预测（多进程）”，单进程与多进程在 ArcGIS 与 ArcGIS Pro 下均可使用。由于 ArcGIS 下 Python 为 32 位，多进程极易出现“内存错误”。

【功能】

此工具根据上一步得到的最优参数模型对研究区滑坡易发性图进行预测。得到的易发性图坐标系与输入的待预测影像一致。

【输入】

- ① 待预测影像路径：必填。选择研究区多通道待预测栅格数据所在目录。
- ② 研究区矢量边界：必填。用于裁剪生成的易发性图，使其保证结果为研究区范围。
- ③ 模型路径：必填。上一步得到的最优模型保存文件。
- ④ 易发性图保存路径：必填。生成的易发性图保存文件。
- ⑤ 数据集行数：必填。数据集样本的行数。
- ⑥ 数据集列数：必填。数据集样本的列数。
- ⑦ （多进程参数）pythonw.exe 所在路径：必填。pythonw.exe 所在路径。ArcGIS

一般为“C:\Python27\ArcGIS10.8”。ArcGIS Pro 一般为“ArcGIS Pro 安装路径\bin\Python\envs\arcgispro-py3”。

【输出】

研究区滑坡易发性图。



图 13 易发性图预测（单进程）



图 14 易发性图预测（多进程）

三、数据准备与案例数据说明

【\Data】文件夹为配套的案例原始数据，数据说明见表 1：

表 1 数据说明

数据准备	数据格式	案例数据
研究区边界	面矢量 (.shp)	\Data\point\study_range.shp
研究区历史滑坡数据	点矢量 (.shp)	\Data\point\landslide_point.shp
DEM	30m 栅格 (.tif)	\Data\big_factor\dem.tif
断层	线矢量 (.shp)	\Data\big_factor\faults.shp
岩性	30m 栅格 (.tif)	\Data\big_factor\lithology.tif
道路	线矢量 (.shp)	\Data\big_factor\roads.shp
河流	线矢量 (.shp)	\Data\big_factor\rivers.shp
NDVI	30m 栅格 (.tif)	\Data\big_factor\NDVI.tif

逐月降雨量

NC4 文件 (.nc4)

\Data\big_factor\rainfall*.nc4

注：所有数据均需投影至 UTM 投影坐标系下。

【\Case】文件夹为使用工具箱生成的论文数据，本手册所有案例数据以 ArcGIS 软件中生成为例，数据说明如下：

- ① \Case\big_factor 文件夹：存放原始数据及未批量裁剪的各影响因子数据。
- ② \Case\point 文件夹：存放研究区矢量数据及滑坡与非滑坡点矢量文件。
- ③ \Case\factors 文件夹：存放裁剪后的研究区各因子层数据。
- ④ \Case\dataset 文件夹：存放未进行因子筛选生成的块数据集，包括滑坡 (\landslide) 与非滑坡样本 (\non-landslide)。
- ⑤ \Case\IGR_dataset 文件夹：存放经过因子筛选后生成的块数据集，包括滑坡 (\landslide) 与非滑坡样本 (\non-landslide)。
- ⑥ \Case\model 文件夹：存放生成的不同参数得到的模型及模型评价结果。
- ⑦ \Case\predict 文件夹：存放生成的待预测影像及预测得到的易发性图。