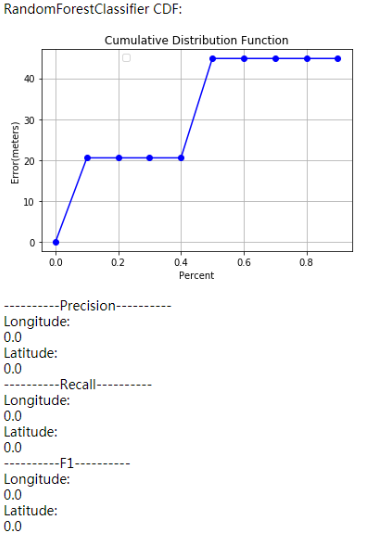
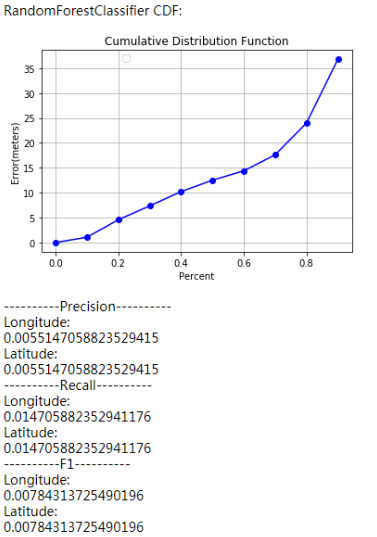
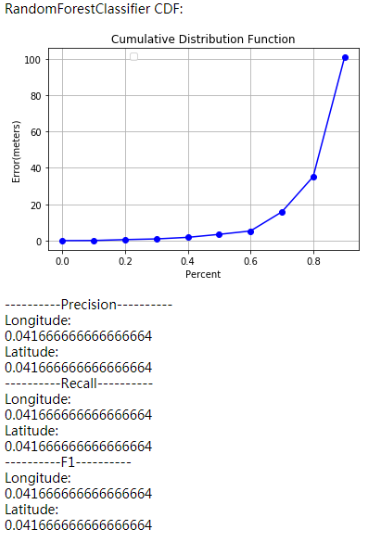
## 代码运行

[点击查看原图](c.png)

（图片较大，请浏览原图）

### 讨论分析

1. 本题通过对数据按主基站进行分组，并去除栅格分类，缩小数据点的所在范围来增加样本间的差异性，来提升预测的效果。
2. 本题仅使用随机森林模型，由a的分析可知是一种十分合适的选择。
3. 以下为部分结果截图：



* 1. 第一张图属于较为普遍的模式，从中可见，由于将数据进行了分组，缩小了区域范围，预测误差降到了100m（即5个栅格宽度）以内，而中位误差则大幅度地降到了10m以内（小于1个栅格宽度），由此可知，此方法对于单个基站精度明显好于a中的方法。
  2. 第二张图为其中较好的结果，其最大误差控制在50m（即2.5个栅格宽度）以内，而中位误差稍大，但依然在20m（即1个栅格宽度）以内。和前者相比虽然图线略优，但是在之后计算topk的过程中实际相差无几。
  3. 第三张也是较为常见的模式之一，其曲线呈现不自然的突变状，观察其数据集可知，其数据量过少，表现在图线中即是一个点可能直接对应一个相邻数据点的误差，为阶跃状。在之后计算topk的过程中实际并不影响，数据量的多少仅影响数据集合并之后的效果，详情可见之后几个小题的报告。

### 性能比较

此处性能比较将和a做对比，比较单组和多组训练之间的耗时差异，总数据量保持一致，取5个样本，总分组为43组，计算平均时间的43倍，多组多次训练的总耗时大于单组训练一次的2倍。