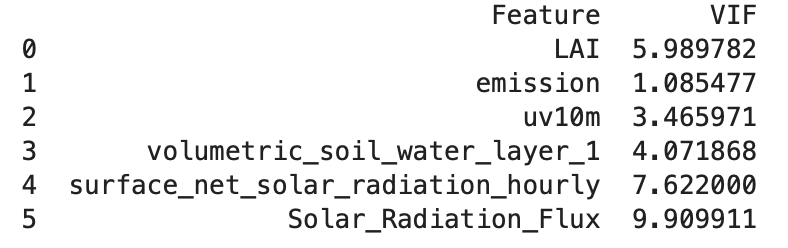
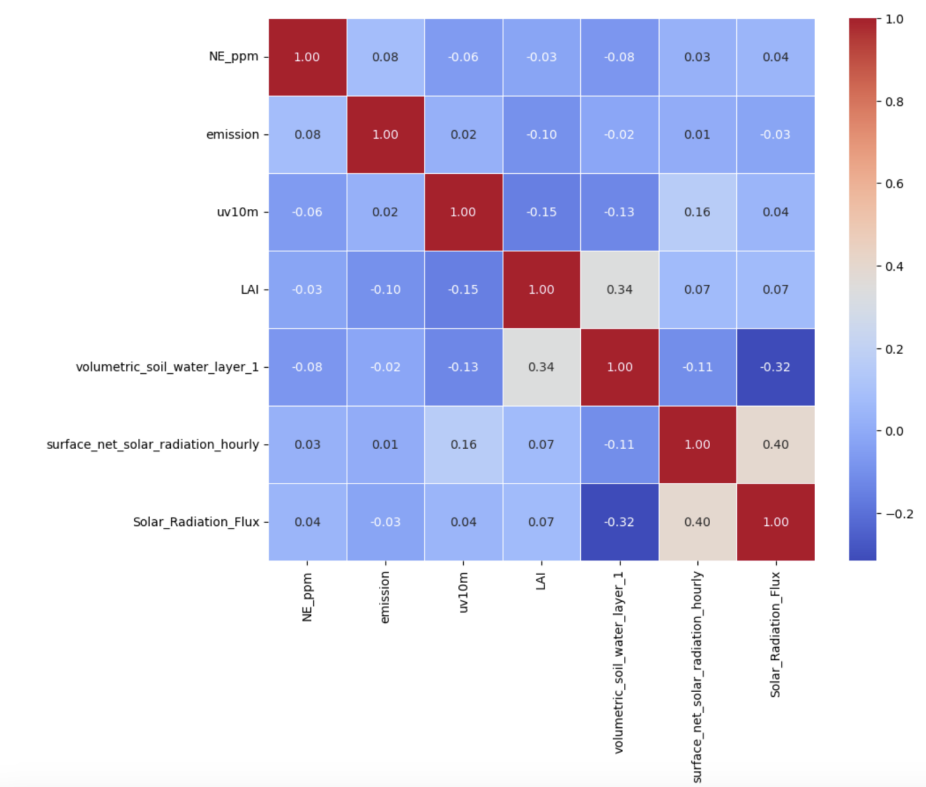
**关于新数据gtwr过程中的疑问**

Chenya 12.7.2024

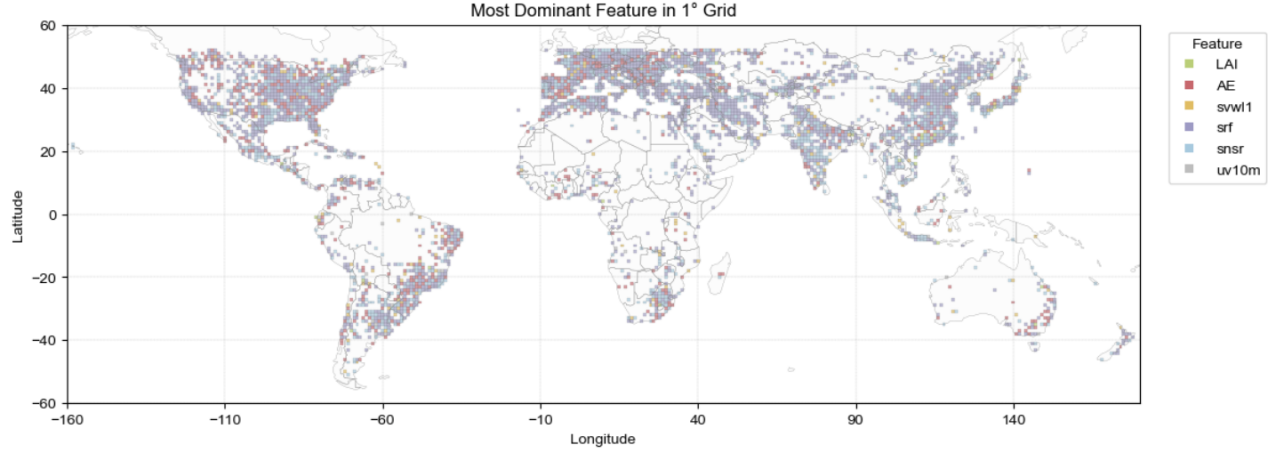
1. 新数据的srf和snsr均表现出大于7.5的vif，比旧数据更严重，表现出高度的共线性，是否应该删掉一个？

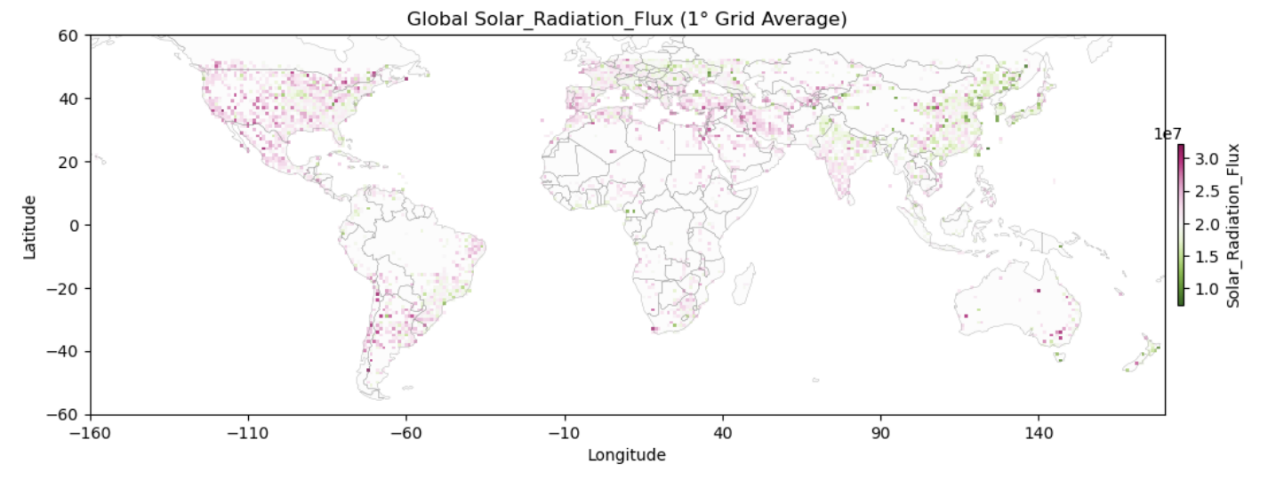


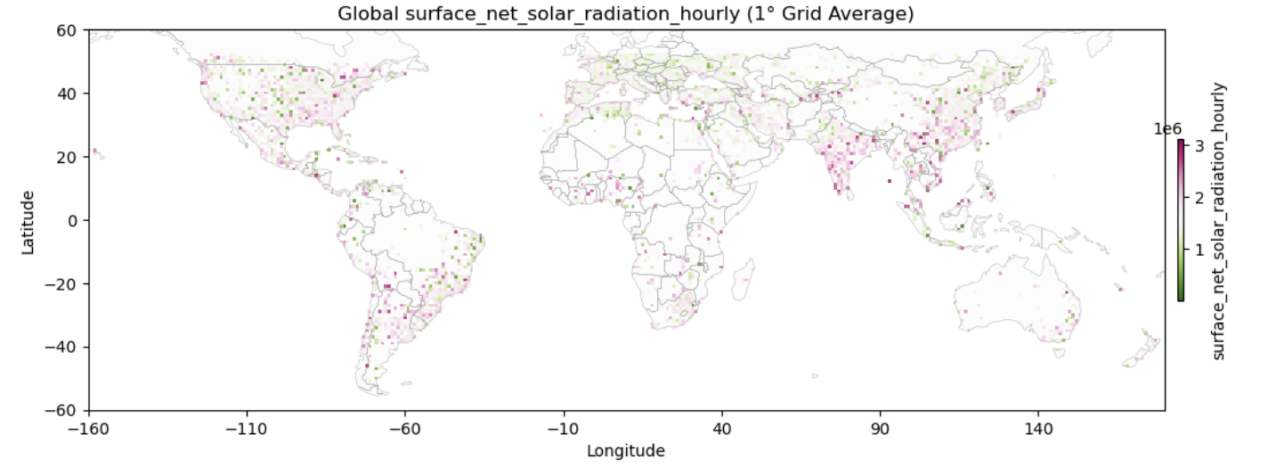


1. 同时，在讨论dominant features时，大部分地区都是srf和snsr （蓝色/紫色）。这两个数据在概念上有一定重叠的部分，并且在较大的尺度保持一致，例如，在美国/印度整体srf都很高，没有区域差异，如下图2/3。辐射总数是与当地经纬度、地形地势非常相关的一组数据，很难受到人类活动影响，特别是建成区已经为不透水面的情况下，从分析的角度来说，判断出大部分城市最重要的是辐射后，可能无法对城市政策法规作出什么影响。基于这个逻辑，我感觉两个radiation的变量选择留下一个就可以，特别是我们研究的是大尺度、长时段的变化，很多细节时间的信息已经被平滑掉了，我感觉保留srf比较好。

后面也可以进一步研究srf的季节性，影响用电等等。瞬时的天气变化感觉对ne的影响和解释均有限，且和srf重合了一部分解释。







1. 另外就是，我重新思考了一下我们的数据，在进行天气因子等的补充后，天气因子是hourly分辨率，ae为daily，他们已经不适合分析细微时间尺度的事件，我们以second为timestamp单位，会使模型提高复杂度，过于关注单点的数据，（相当于引入了一些噪音），是否不太合适？是不是从hour为单位开始做modeling才是比较合适的尺度？
2. 风现在有风速的数据，要不要加方向的数据？新加一个标量，结合风速可以判断不同季节污染物传输的方向，考虑季风（主要风向等），促进regional减少碳排合作。