

写给贩卖人员的一段话:

长久以来,我一直受诸位同行的盗卖,有人盗卖说明我们的质量取得了大家的认可。但是,一直以来都存在高价贩卖【99元的资料售卖到900元、2900元】,售后资料后不进行答疑【购买资料后不做任何售后,我们的售后群不禁言】,这不仅欺骗了大量的上帝,也对我们的口碑产生了负面影响。因此,我们团队与面包多平台进行协商,确定了一种全新的保护方式。

对于我们发布的资料,平台提取作品特征码(作品创建时间以及作品指纹MD5码),同时我们还对每个付费之后的作品插入完全不同的标记【每个用户插入的标记完全不一样,同一作品、不用用户的区别只有一个字符,不会影响正常阅读使用,仅为防止贩卖】,这会使得每一个用户拿到的作品特征码均不相同。我们将根据您贩卖的版本提交平台,平台将直接锁定您的个人账号。并利用已经取得公证处盖章/最高法院承认/国家授时中心&公证云可查的证明文件和信息对您发起相关诉讼。

如果,您想售卖我们的资料,可以与微信**bzdsxjm521**联系,我们会直接给您一个合适的分销链接,保证分成合适。但我们不能接受您高价转卖,资料不进行售后。

最后,各位生活愉快!!!!!!

注:

本次竞赛作品word文档已经取得公证处盖章/最高法院承认/国家授时中心&公证云可查的证明文件和信息

最高法院 查询地址: <https://sf.court.gov.cn/pages/verification/submitverification>

国家授时中心 查询地址: <https://ttas.ntsc.ac.cn/>

公证云 查询地址: https://www.ezcun.com/czcx/czcx_evid.html

注: 该作品参赛粘贴使用, 主要针对不劳而获高价转卖的二手贩子

必读!!! 为了防伪, 做出以下防盗措施

1、目前论文中所有图片均有水印, 结束前一天将公布无水印版本。

2、在写作过程中, 本文求解过程存在明显的模型逻辑问题, 该问题会在

BZD 官方认可的渠道面包多进行购买加入的售后群, 会在售后群教大家如何修改明显的模型问题。如果不改会对论文整体质量有影响。

注: 在其他渠道购买, 写出论文存在漏洞扣分等问题, 概不负责

BZD 数模社官方认可链接:

<https://mbd.pub/o/author-bGWSnGprag>

主页截图, 如下所示

BZD数模社

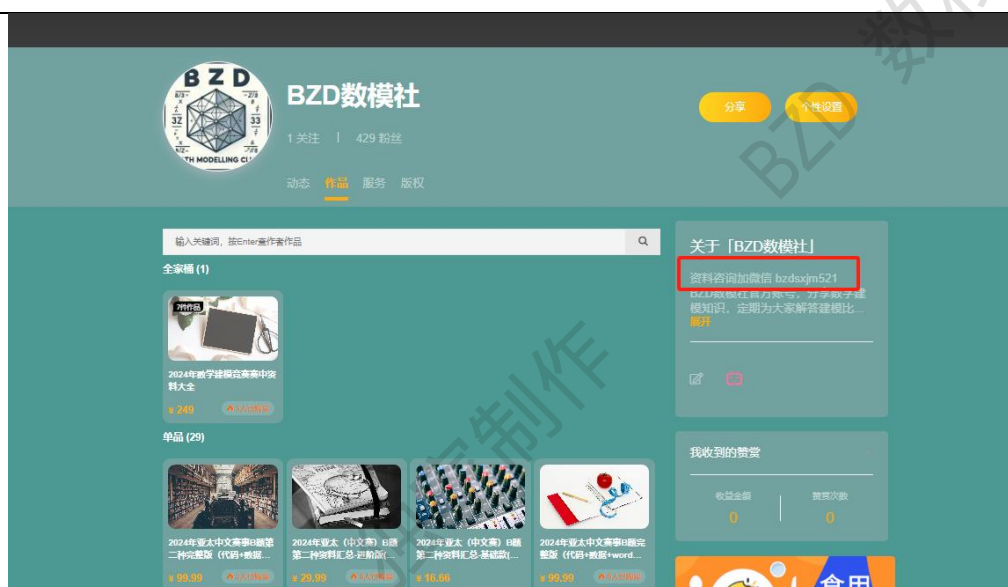
2024.09.22

倒卖必究

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接

<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>



使用须知： 该论文完全按照提交格式进行写作，很多人都会看到这一版本成品论文，直接提交一定会查重不过关。成品论文是按着之前的解题思路写出的，如果不对外展示，在普通建模竞赛是可以直接提交 100%获奖的文章，保二冲一的水平。进行展示的目的，是为了辅助大家写论文。也让大家心里有个数大概获奖论文是个什么层次。【写的比好就是一、二等奖，比差就是三等奖、优秀奖】

组委会要求重要格式规范如下所示，大家可以根据要求，以及 1 资料中提供的优秀论文资料进行修改。以下是竞赛论文要求，请认真阅读！！

- 1、第一页还有论文标题、摘要和关键词（无需译成英文），并从此页开始编写页码；页码必须位于每页页脚中部，用阿拉伯数字从“1”开始连续编号。
- 2、论文从第二页开始是论文正文；正文之后是参考文献和论文附录（页数不限）。附录需要填写一个表格，交代程序代码名称等信息。
- 3、引用别人的成果或其他公开的资料(包括网上查到的资料) 必须按照规定的参考文献的表述方式在正文引用处和参考文献中均明确列出。正文引用处用方括号标示参考文献的编号，如[1][3]等；引用书籍还必须指出页码。参考文献按正文中的引用次序列出。

使用、借鉴、抄写该论文是否违规：只要不超过查重率的 20%，就不能算作违规。该论文属于公开发表的网上资料，进行借鉴、参考不能算作违规，切记不要抄的太过分，查重率一定要低于 20%就可以。下图为最新的数模竞赛规定

第七条 所有引用他人或公开资料(包括网上资料)的成果必须按照科技论文的规范列出参考文献，并在正文引用处予以标注。

下面为成品论文 正文

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接

<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群：689964173 数模交流群 2 群：275032074
国赛助攻资料通知群：928949323

BZD数模社

2024.09.22

倒卖必究

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接
<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>



中国研究生创新实践系列大赛
“华为杯”第二十一届中国研究生
数学建模竞赛

学 校	BZD
	BZD
参赛队号	
	1.B
队员姓名	2.Z
	3.D

BZD数模社

2024.09.22

倒卖必究

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323

题 目： 大数据驱动的地理综合问题

摘 要：

地理系统是自然、人文多要素综合作用的复杂巨系统，地理学家常用地理综合的方式对地理系统进行主导特征的表达，本文利用大数据的手段对地理系统进行综合，探索全球气候变化下中国地理环境的演化。

针对问题一，本文首先对数据进行清洗，替换一些取值较大或较小的特殊值，并利用 $3-\sigma$ 准则确定一些离群点，然后使用数字、图表等方式，对原始数据进行定量总结、概括，得出了一些降水量、土地利用/土地覆被面积两个变量的在 1990 至 2020 年间中国范围内的时空演化特征。

针对问题二，首先利用逻辑回归模型量化地形-气候相互作用在极端天气形成过程中的作用，再用格兰杰因果检验和斯皮尔曼相关系数加以检验，确定它们之间的相互作用，验证了本文模型建立的有效性，为后文预测的准确性奠定基础。

针对问题三，首先对题目中提到的自变量进行量化，建立逻辑回归模型，再利用移动平均线模型和 LSTM 神经网络进行预测，将预测的数据代入前面建立的逻辑回归模型进行降水量的预测，利用不同的成灾临界值可确定不同的防范政策。若需要推广至多分类问题，为不同地区制定不同政策，也即对全国所有城市分类出应对暴雨灾害能力最为脆弱、较为脆弱、一般、不脆弱等类的话，只需将本文中的成灾临界值调为阶梯型的函数即可，增强模型的泛化能力。

针对问题四，将数据集 3 的降水量、数据集 5 的人口数量、数据集 6 的 GDP 数据作为中国土地利用变化的影响因素，重复问题一的描述性统计，与问题二、三的建模进行分析，描述中国土地利用变化的特征与结构，并从准确性和有用性两个方面总结解释本文所建立的模型与获得的结果。

关键词：逻辑回归；LSTM；大数据可视化；格兰杰因果检验；斯皮尔曼相关系数

一、问题重述

1.1 问题背景

地理系统是自然、人文多要素综合作用的复杂巨系统，地理学家常用地理综合的方式对地理系统进行主导特征的表达。如以三大阶梯概括中国的地形特征，以秦岭—淮河一线和其它地理区划的方式揭示中国气温、降水、植被、土壤及生态环境在水平和垂直方向上的地带性与非地带性规律，利用胡焕庸线、T 型开发结构等描绘我国人口、社会和经济发展的总体格局。这些方法早期以宏观结构和定性分析为主体，对我国生态保护、社会经济发展和国家安全保障起到了巨大的支撑作用。伴随着对地观测体系的快速发展，当前已经积累了巨量的对地观测数据。如何利用大数据的手段对地理系统进行综合，探索全球气候变化下中国地理环境的演化，是当前地球科学研究的关键问题。

1.1 问题回顾

问题 1：在众多描述地理环境的变量中，一些简单的指标背后蕴藏了深厚的内涵，对人类的生存发展具有重大深远的影响，如大气中二氧化碳的浓度、全球年平均气温等。降水量是一个连续变化的变量，而土地利用/土地覆被类型则是一个存在突变和离散分布的变量。同时，它们都具有时空分布不均匀的特征。请从附件数据中选取相关数据集，为这两个变量分别构建一套描述性统计方法，用 1 到 3 个较为简洁的统计指标或统计图表，对这两个变量在 1990~2020 年间中国范围内的时空演化特征进行描述和总结。

问题 2：近年来，以暴雨为代表的极端天气事件对人类的生产生活造成了越来越难以忽视的影响。请结合附件中所给的数据，建立数学模型，说明地形-气候相互作用在极端天气形成过程中的作用。

问题 3：降雨、地形和土地利用对于暴雨等极端天气灾害的形成都具有不可忽视的影响。这其中，降雨的时空变异性和不可控性都最强；土地利用作为自然条件和人类活动的综合结果，虽然也随时空演化，但具有一定可控性；地形是最为稳定、不易改变的因素。

请考虑第 2 问所反映的从“暴雨”到“灾害”中上述三方面因素的角色及其交互作用，确

定暴雨成灾的临界条件；并结合第 1 问中降雨量和土地利用/土地覆被变化的历史时空演化特征，对 2025 至 2035 年间中国境内应对暴雨灾害能力最为脆弱的地区进行预测。请以地图的形式呈现你们的预测结果。

问题 4：在中国级别的尺度上，描述自然地理特征的地形可以概括为“三级阶梯”，而降水中具有标志性意义的“800mm 等降水量线”则与区分我国南北方的“秦岭—淮河”一线大体重合；描述人文地理特征的人口分布及其社会经济活动总量等指标，则被由连接黑龙江黑河与云南腾冲的“胡焕庸线”清晰地划分成东密西疏的两部分。那么，对于自然地理和人文地理交汇点的土地利用/土地覆被情况，结合其在前三问中描述、估计和预测任务中的“特性”，利用地理大数据，建立相应的数学模型，对数据进行简化和综合，描述中国土地利用变化的特征与结构。从准确性和有用性两个方面解释验证你们的总结。

二、问题分析

2.1 问题一分析

针对问题一，本文首先对数据进行清洗，替换一些取值较大或较小的特殊值，然后使用数字、图表等方式，对原始数据进行定量总结、概括，得出了一些降水量、土地利用/土地覆被面积两个变量的在 1990 至 2020 年间中国范围内的时空演化特征。

2.2 问题二分析

针对问题二，首先利用逻辑回归量化地形-气候相互作用在极端天气形成过程中的作用，再用格兰杰因果检验和斯皮尔曼相关系数加以检验，确定它们之间的相互作用，验证了本文模型建立的有效性，为后文预测的准确性奠定基础。

2.3 问题三分析

针对问题三，首先对题目中提到的自变量进行量化，建立逻辑回归模型，再利用移动平均线模型和 LSTM 神经网络进行预测，将预测的数据代入前面建立的逻辑回归模型进行

降水量的预测，利用不同的成灾临界值可确定不同的防范政策。若需要推广至多分类问题，为不同地区制定不同政策，也即对全国所有城市分类出应对暴雨灾害能力最为脆弱、较为脆弱、一般、不脆弱等类的话，只需将本文中的成灾临界值调为阶梯型的函数即可，增强模型的泛化能力。

2.3 问题四分析

针对问题四，将数据集 3 的降水量、数据集 5 的人口数量、数据集 6 的 GDP 数据作为中国土地利用变化的影响因素，重复问题一的描述性统计，与问题二、三的建模进行分析，描述中国土地利用变化的特征与结构，并从准确性和有用性两个方面总结解释本文所建立的模型与获得的结果。

三、模型假设

- 1、假设所有使用的气象、地形和土地利用数据都是准确和可靠的。
- 2、假设在研究期间内，中国的地形变化不大，可以认为是稳定的。
- 3、假设在未来预测期间，现有的社会经济发展趋势和政策导向将持续
- 4、在模型中，假设人类活动对土地利用变化的影响可以通过现有数据进行合理估计，并在模型中得到体现。
- 5、在分析暴雨成灾的临界条件时，假设一个地区的灾害风险主要受当地气候、地形和土地利用因素的影响，而较少受到远离地区的影响。
- 6、假设数据记录时不存在漏记错记的情况。

四、符号说明

符号	说明
X_{tij}	Logistic 模型中的自变量
Y_{tij}	Logistic 模型的自变量

Z_{tij}	Logistic 模型的因变量
f	自变量与因变量之间的映射机制
$MAPE$	平均绝对百分比误差
d_i	第 i 个序列误差
ρ	斯皮尔曼相关系数
α	置信水平
h_{t-1}	LSTM 中的隐藏状态
W_i	LSTM 中的可学习参数
σ	标准差

注：这里只列出论文各部分通用符号，个别模型单独使用的符号在首次引用时会进行说明。

五、模型建立与求解

5.1 数据清洗

对于数据集 3，数据集中将中国以外的经纬度上的降水量数据均设为了-99.9，在后续数据处理中，因问题一中需要建立统计指标与统计图表，将-99.9 设为 0，防止这些数据对一些边界地区的降水量指标造成影响。对于人口、GDP 等数据的处理方式相同，将其中的-NAN 或 NAN 替换为 0。

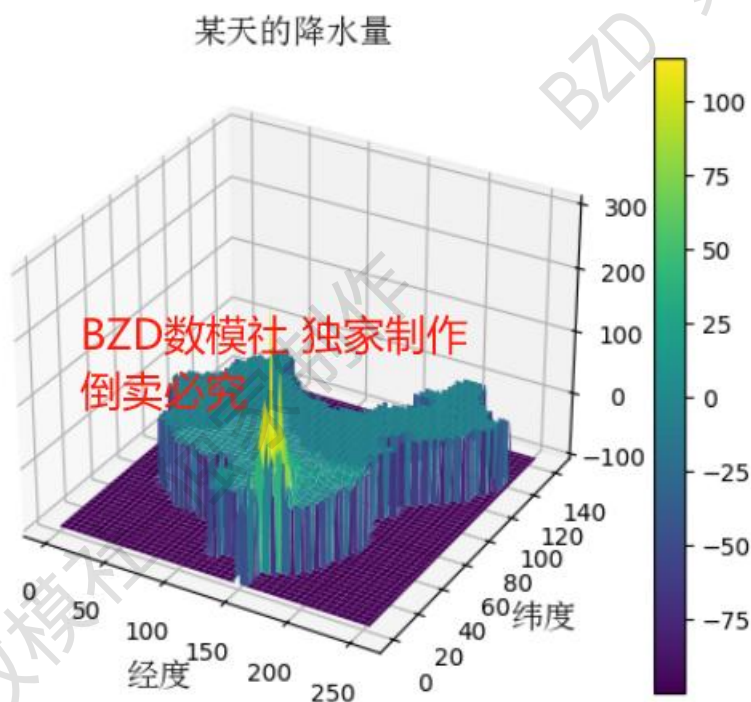


图 1 数据集 3 中取某一天的降水量的可视化

利用 Python 进行编程将负值设为 0，并进行可视化后的结果可见图 2。将 NetCDF 文件中的数据 `daily precipitation` 也即 `pre` 的数据取出转化为矩阵的形式，利用 $3-\sigma$ 准则对异常数据进行分析，此时并不进行剔除，在问题二和三中对“暴雨”和“成灾”界定后再进行剔除。于此同时，对于其他数据集的如 GDP、人口、地形、气温，土地利用和覆盖，以每一年为一个样本，利用 $3-\sigma$ 准则对异常数据进行分析并剔除。

BZD数模社

2024.09.22

倒卖必究

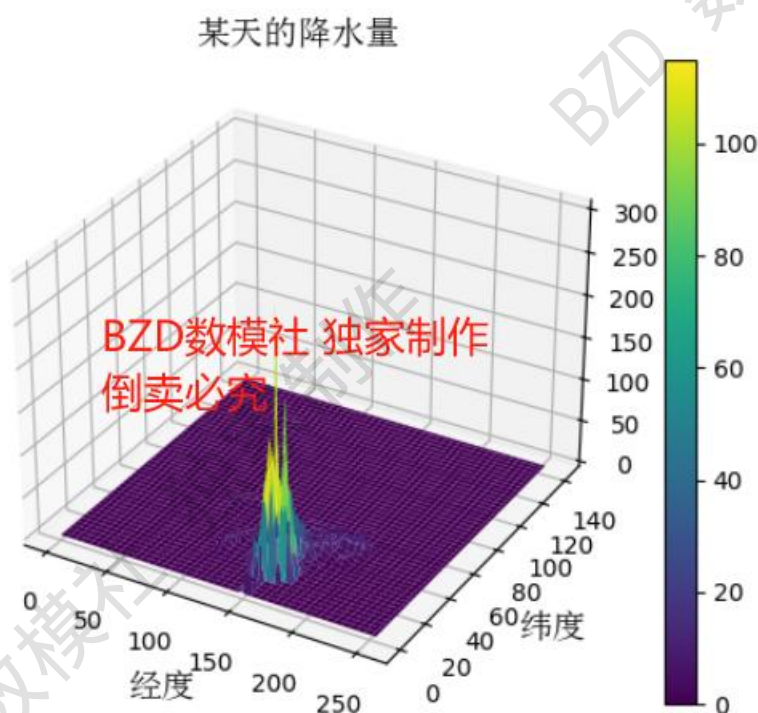


图 2 处理后的数据集 3 的降水量可视化

5.2 问题一模型的建立与求解

5.2.1 问题一模型的建立与求解

首先对这降水量在 1990 至 2020 年间中国范围内的时空演化特征进行描述和总结，首先固定空间分析该变量随时间变化的趋势，

BZD数模社

2024.09.22

倒卖必究

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接

<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

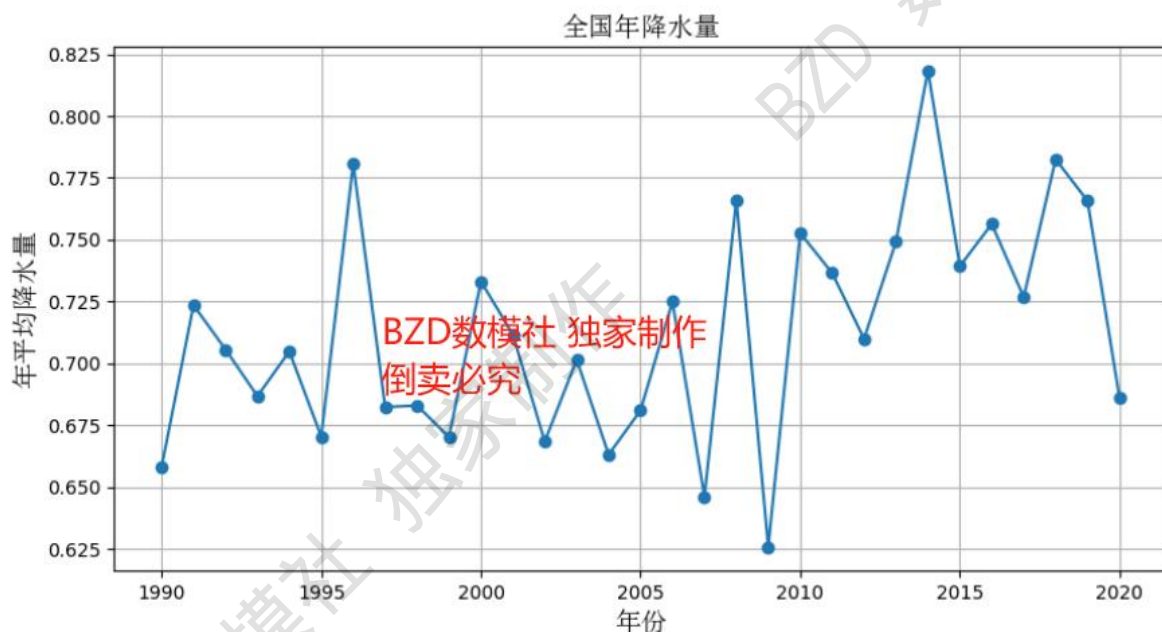


图 3

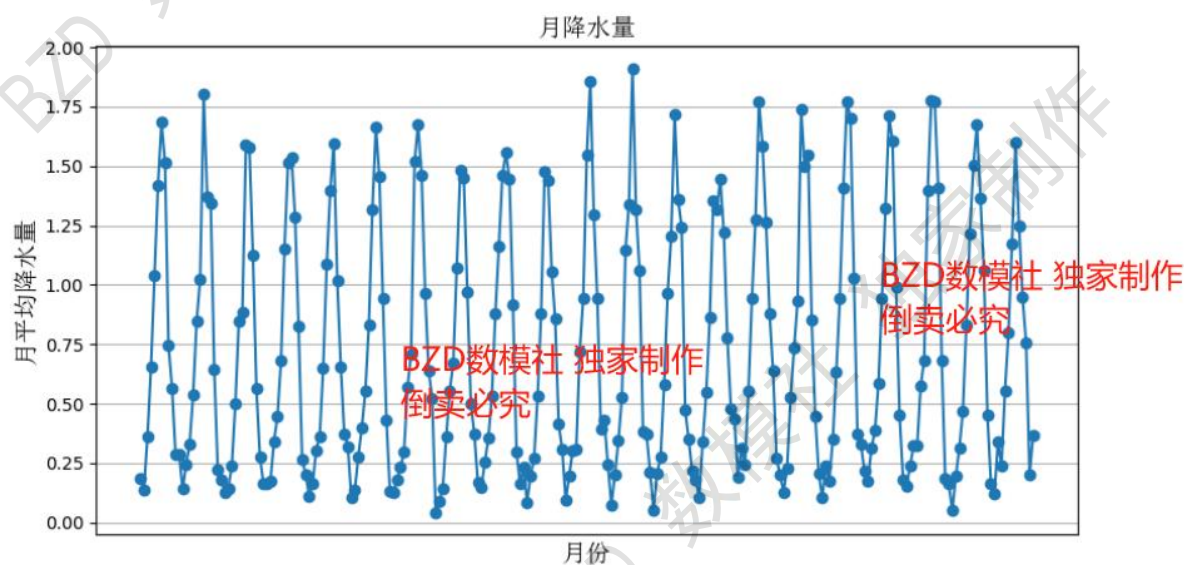


图 4

从波动来看最大值与最小值的差距并不是特别明显，每年基本上都维持在当地的一个平均水平上。但从月降水量的波动来看，全国的月降水量明显呈现出季节趋势，存在着明显的波峰与波谷。（为降重考虑，大家可以自行补充语文描述，分析全国的降雨量随时间的变化趋势）

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群：689964173 数模交流群 2 群：275032074
国赛助攻资料通知群：928949323

2024.09.22

倒卖必究

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接
<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

表 1 固定的地区的经纬度范围

	经度	纬度
山东省	114.8 - 122.7	34.37 - 39.4
北京市	115.4 - 117.5	39.4 - 41.0
西藏自治区	78.3 - 99.1	26.8 - 36.4
吉林省	121.6 - 131.3	40.8 - 46.3
江苏省	116.3 - 121.9	30.8 - 35.1
四川省	97.3 - 108.5	26.0 - 34.3

根据表 1，计算特定地区的年降水量，见下图，从各个省份或城市的降水量波动来看，与年降水量变化时一致的，最大值与最小值的差距并不是特别明显，每年基本上都维持在当地的一个平均水平上。但不同地区的年降水量就截然不同，可以明显的看出南方的省份和城市降水量要明显高于其他地区的城市。



相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接

<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

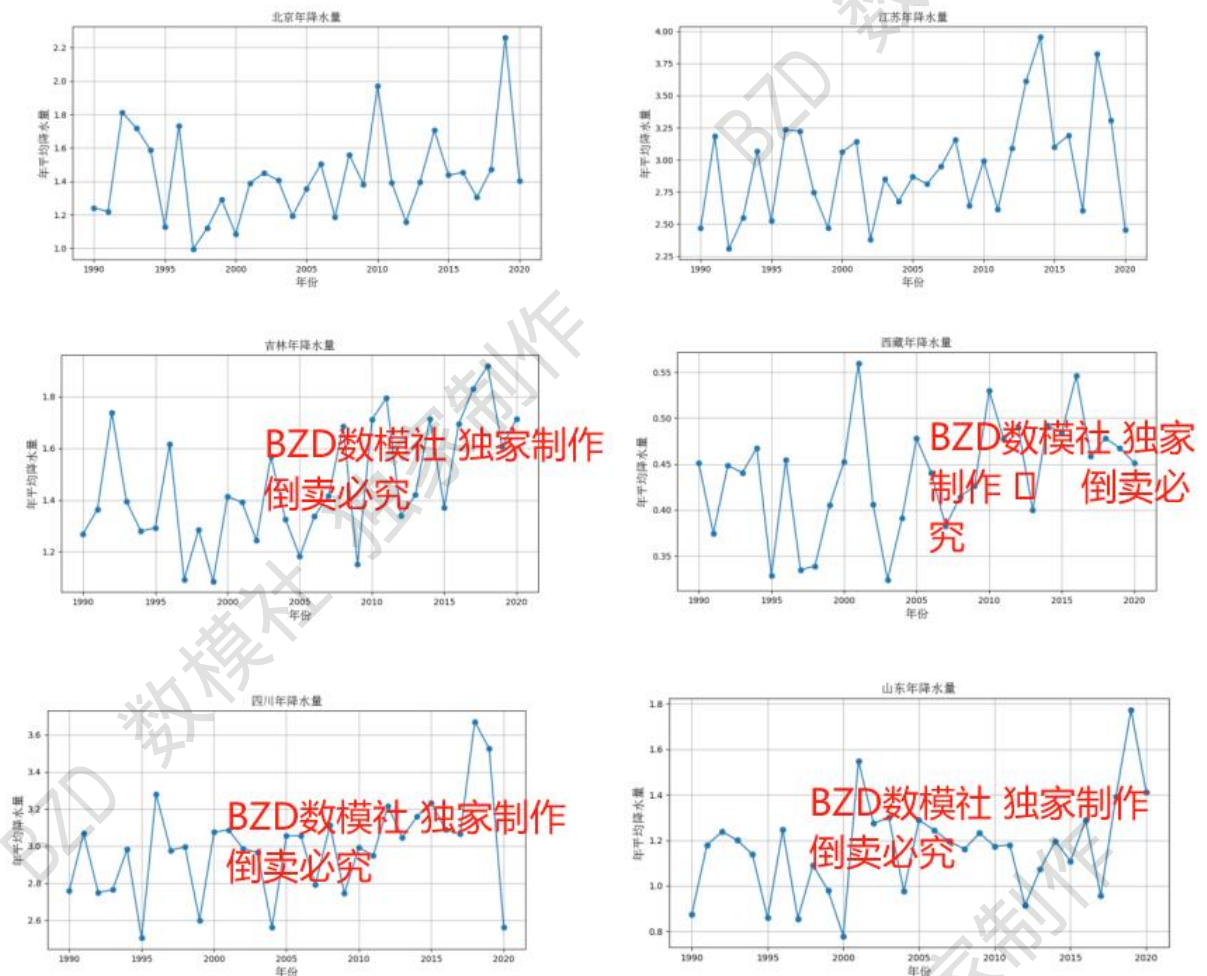


图5

此时，固定时间分析降水量在空间上的演化特征，从时间和空间的分析来看，我们可以得出以下结论：

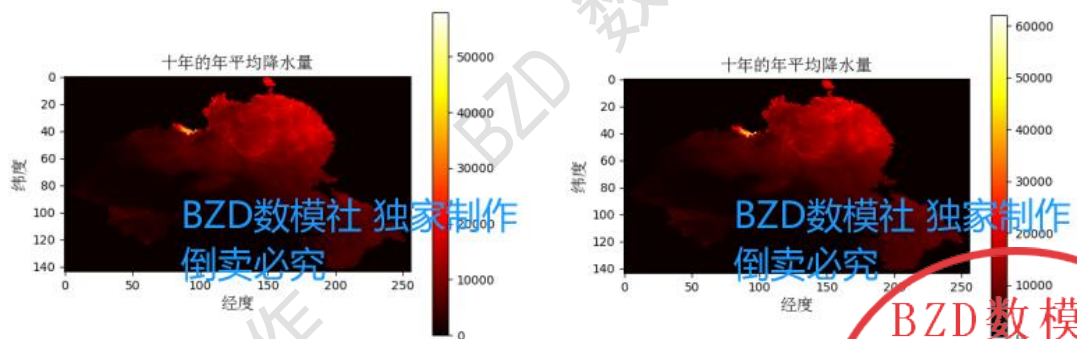


图6 左图为2000年至2010年的平均降水量热力图，右图为2010年至2020年

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群：689964173 数模交流群 2 群：275032074
国赛助攻资料通知群：928949323

2024.09.22

倒卖必究

对于耕地、森林覆被同理分析，其他地区类型不再赘述，演化特征的分析都是相同的，首先固定空间分析该变量随时间演化的趋势，这里本文固定空间为全国，以全国的时间演化趋势分析为例，每十年为一周期，计算十年内的覆盖平均值，结果如下图所示，不论是耕地还是森林覆盖面积，其实从十年的平均值来看，其实变化都不大，说明了土地利用作为自然条件和人类活动的综合结果，虽然也随时空演化，但具有一定可控性。

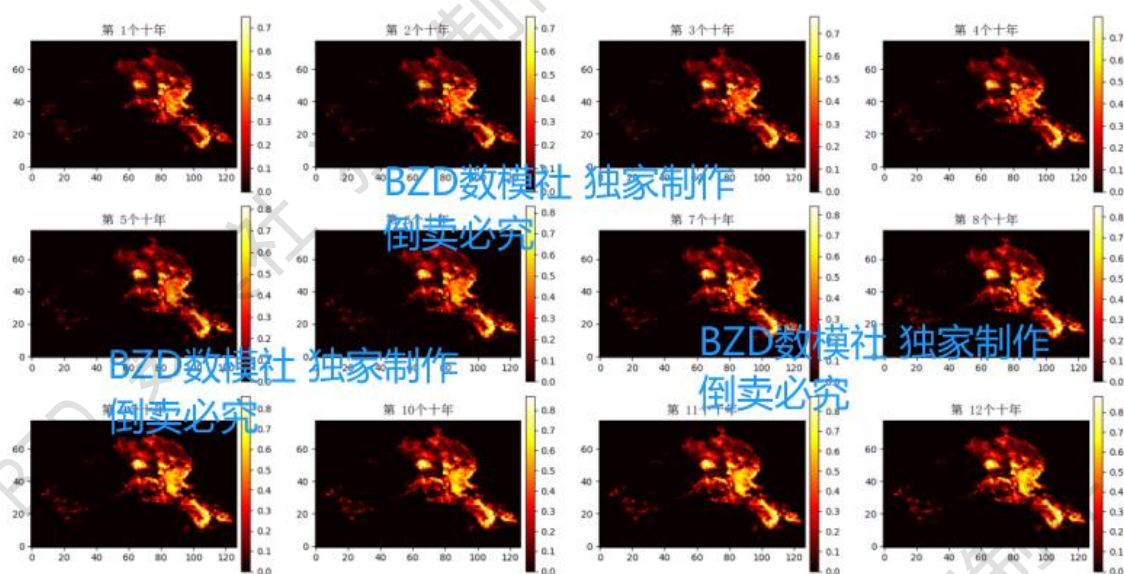


图 7 全国耕地覆盖面积热力图

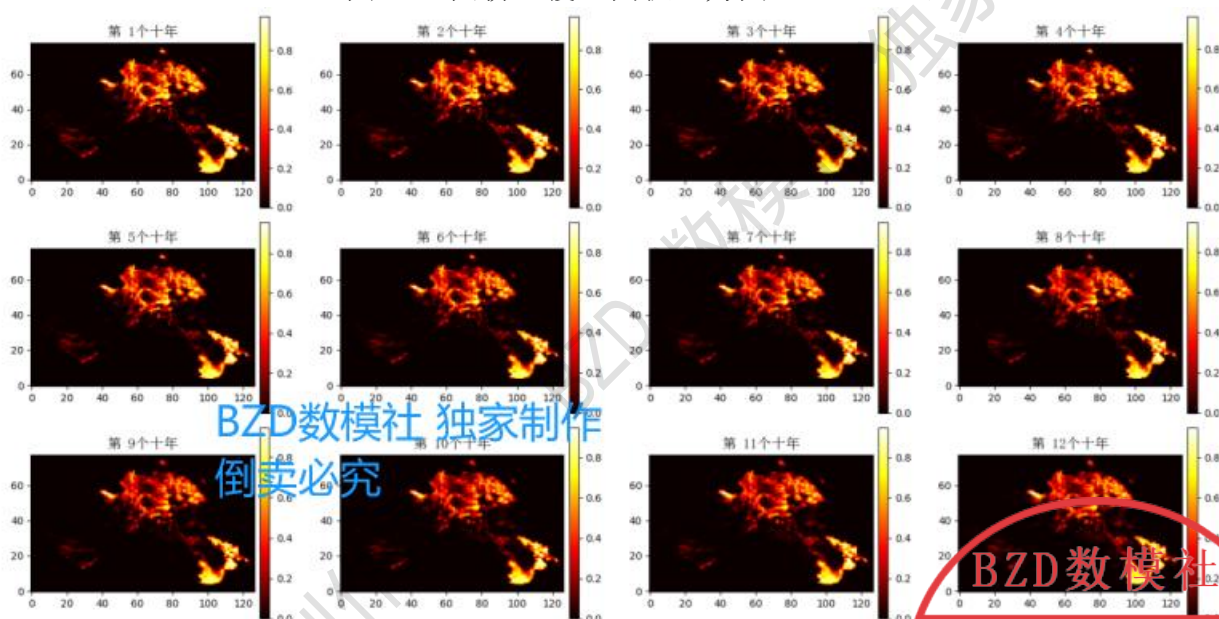


图 8 全国森林覆盖面积热力图

然后固定时间分析耕地覆被在空间上的演化特征，从上图中即可看出中原地区的耕地面积要明显的更多，而中国版图的上方内蒙古附近的森林较多。从时间和空间的分析来看，我们可以得出以下结论：

- 1、年降水量基本上都维持在一个平均水平附近。
- 2、月降水量明显呈现季节趋势，存在着明显的周期性上升和下降趋势，并且存在着周期性的波峰与波谷。
- 3、南方地区的降水量要明显比北方城市的降水量多。
- 4、不论是耕地还是森林覆盖面积，十年的平均值的分布变化都不大，土地利用时空演化具有一定可控性。
- 5、森林覆盖面积与耕地覆盖面积全国分布差异较为明显，不同地区的覆盖面积相差很大。

5.3 问题二模型的建立与求解

5.3.1 地形与气候的量化

对于地形数据，首先对数据集 1 中的 WGS84 地理坐标系的数据进行可视化，可见图 7，然后对所有数据进行归一化，防止量纲不同影响模型的准确程度。对于气候同理进行归一化，气候以每年为单位做归一化，也即对每一个文件夹下面的所有图片的所有像素做归一化。

对于数据集 1 中的地形数据，形状为 $x_{ij}, i = 0, 1, \dots, 5384, j = 0, 1, \dots, 10891$ ，对于数据集 2 中的气温数据，形状为 $y_{tij}, i = 0, 1, \dots, 399, j = 0, 1, \dots, 699$ ，对于数据集 3 中的降水量数据，形状为 $z_{tij}, i = 0, 1, \dots, 143, j = 0, 1, \dots, 255$ 。问题二要求我们建立模型来量化地形与气候相互作用在极端天气（暴雨）形成过程中的作用，也即量化出下述 f 的表达式：

$$f(x_{ij} + y_{tij}) = z_{ij}, \quad \forall t, i, j$$

其中将量化完的地形数据沿着时间维度复制，将时间维度与降水量的时间维度对齐，

考虑到时间久远对现阶段地形与气候相互作用在极端天气（暴雨）形成过程中的作用的贡献并不明显，本文仅采用 2008 年至 2018 年 1 月 1 日至 12 月 31 日的每一日的数据。另外在空间维度上的对齐采用大图像取平均值减小像素点的个数从而对齐小图像，也即最终所有数据的空间维度均为（144，256），若希望达到更好的精度，可以用小图像对齐大图像，也即最终所有数据的空间维度均为（5385，10892），本文采用效率较高的算法来求解下述量化 f 的逻辑回归模型，高精度的求解是同理的，因此本文选择了前者方法。

5.3.2 Logistic 回归模型的建立

逻辑回归是处理二分类问题的强大工具，在本文中，我们利用数据构建了逻辑回归模型，并对模型进行了训练，采用梯度下降的迭代方法，利用 SPSS 对逻辑回归模型中的参数进行拟合。

在模型构建过程中，逻辑回归通过估计一个事件的对数几率与一个或多个自变量（特征）之间的线性关系。在本研究中，模型的目标是预测极端事件（暴雨）是否发生，模型的输出是一个介于 0 和 1 之间的概率值，表示暴雨发生的可能性。通过设定一个阈值，本文选取的为 0.7，同时也将降水量排序后的数据的 70% 设为正常降雨，后 30% 设为暴雨，我们可以将概率值转换为二元类别，从而完成分类。由于对暴雨的界定不同，可以灵活修改模型中的阈值参数，以灵活适应现实标准，模型的灵敏度分析中，本文也将指出这一点。

逻辑回归模型的基本形式可以表示为：

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$

其中， $P(Y = 1|X)$ 是给定自变量 X 时因变量 Y 等于 1 的概率， e 是自然对数的底数， $\beta_0, \beta_1, \dots, \beta_k$ 是模型参数。

利用 SPSS 进行求解后，得到系数如下所示：

表 2

参数	迭代到的局部最优解
β_0	0.1246
β_1	3.4622

β_2

-1.8710

因此地形-气候相互作用在极端天气形成过程中的作用可以由逻辑回归确定的模型所量化。

5.3.3 Grander 因果检验

利用 Grander 因果检验来检验地形的海拔高度、温度与极端天气发生的与是否具有因果关系，该因果关系并不是实际生活中的因果关系，而是来说明海拔高度与温度对极端天气有没有预测效果，仅此一项检验模型来说明极端天气发生与否与温室效应有关显然说服力不高，因此后面对海拔高度、温度与极端天气发生与否做斯皮尔曼相关性分析来进一步确定两者的关系。Grander 因果检验的实现流程框图如下所示：

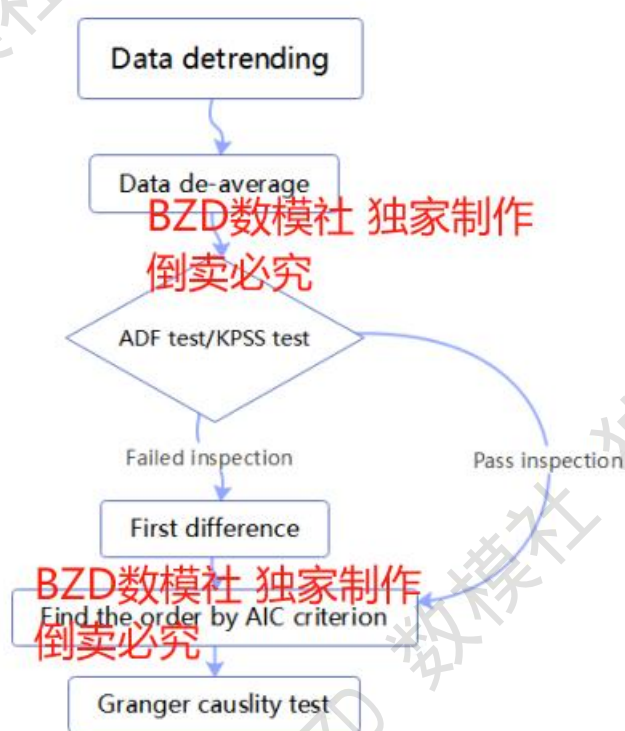


图 9 grander 因果检验流程框图

利用 EViews 对待检验的数据进行分析，在进行 ADF test 时，三种数据均未通过检验，但在二阶差分后平稳性良好，最终 granger 因果检验的结果如下：

表 3 地形的海拔高度和极端天气事件的 grander 检验

Null Hypothesis:	Obs	F-Statistic	Prob.
SERIES03 does not Granger Cause SERIES02	21	1.36738	0.3320
SERIES02 does not Granger Cause SERIES03		4.10927	0.0350

表 4 温度和极端天气事件的 grander 检验

Null Hypothesis:	Obs	F-Statistic	Prob.
SERIES06 does not Granger Cause SERIES03	21	3.37320	0.0480
SERIES03 does not Granger Cause SERIES06		0.21501	0.9614

根据表中结果可以看出：两次检验的 P 值均小于 0.05，拒绝原假设，地形的海拔高度与温度均为极端天气发生与否的 Granger Cause. 下面进行斯皮尔曼相关性进一步验证此关系。

5.3.4 斯皮尔曼相关性系数分析模型

斯皮尔曼相关系数法是一种良好的用于评价两变量间相关性的统计学方法，它具有的最显著特征为不需考察变量的样本规模或总体的分布特性，只需要两组数据观测值的等级评定数据是成对的。以 X_i 和 Y_i 对应相应第 i ($1 \leq i \leq n$) 个元素，对 X 和 Y 按照同种升序或降序方式排列，得到新变量序列 x 和 y ，相应地， x_i 和 y_i 分别对应第 i ($1 \leq i \leq n$) 个元素，以 $d_i = x_i - y_i$ 为元素差分集合，计算随机变量 X 和 Y 之间斯皮尔曼相关系数公式如下：

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (18)$$

其中，分母是与变量序列长度有关的常数，分子是序列误差之和，体现变量间相关差异。系数 ρ 取值范围为 $[-1, 1]$ ，反映两变量之间变化趋势的方向以及程度，其值为 0 则表示变量间不相关，负值则表示变量间呈负相关关系，正值则表示变量间呈正相关关系；该值绝对值越大，表示两变量间相关性越强。

利用 SPSS 分别分析地形的海拔高度、温度与极端天气发生与否的相关性，结果如下：

表 5 地形的海拔高度和极端天气发生与否相关性

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接
<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

相关性			极端天气发生与否	海拔
斯皮尔曼 Rho	极端天气发生与否	相关系数	1	.556**
		Sig. (双尾)	.	0.003
		N	27	27
	海拔	相关系数	.556**	1
		Sig. (双尾)	0.003	.
		N	27	27

表 6 温度和极端天气发生与否相关性

相关性			极端天气发生与否	温度
斯皮尔曼 Rho	极端天气发生与否	相关系数	1	-.538**
		Sig. (双尾)	.	0.004
		N	27	27
	温度	相关系数	-.538**	1
		Sig. (双尾)	0.004	.
		N	27	27

** 在 0.01 级别（双尾），相关性显著。

从表中结果可以发现：地形的海拔高度与极端天气发生与否有显著的正相关关系，温度与极端天气事件发生频次有较显著的负相关关系。

5.3.3 结果分析

因此综合上述两个模型的结果来看，（这段叙述是空缺，可以自行补充，也可以等待后续补充资料，预计 9 月 22 日晚之前，所有的结果语文叙述补充完整）

5.4 问题三模型的建立与求解

5.4.1 建立降雨、地形和土地利用对于暴雨的作用的 Logistic 模型

数据处理方式与问题二完全相同，不再赘述，其中降雨量作为因变量，因此自变量雨将由第二问中已分析的地形与气候来进行量化，因此逻辑回归模型如下所示：

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \frac{1}{1 + e^{-(0.1246 + 3.4622X_1 - 1.871X_2)}} + \beta_2 X_1 + \beta_3 X_3)}}$$

其中X₁为地形的海拔高度，X₂为温度，X₃为土地利用，土地利用为五种土地分布的平均值。



另外其中的阈值此时设为 0.9. 利用 SPSS 对逻辑回归模型中的参数进行拟合。利用 SPSS 进行求解后，得到系数如下所示：

表 7

参数	迭代到的局部最优解
β_0	0.0945
β_1	8.7682
β_2	-2.3780
β_3	5.5984

因此极端天气形成过程已由逻辑回归确定的模型所量化。

5.4.2 建立自变量预测模型—LSTM 循环神经网络

自变量由地形的海拔高度、温度、土地利用构成，其中海拔高度我们假设未来十年不会发生很大的变化，且由于我们已知的数据只有一张地形图，因此地形的数据沿用已知数据，不再进行预测，另外土地利用构成虽然也随时空演化，但具有一定可控性，因此采用简单的移动平均线模型进行预测

在时间序列模型中，简单移动平均法与加权移动平均法均比较适合没有明显趋势变动的的时间序列分析，而当时间序列有直线增加或减少趋势时，此两种方法预测会有滞后偏差影响。趋势移动平均法是对该两种方法的修正，利用二次移动平均，借助移动平均滞后偏差规律，通过两次移动平均适应数据集特点，从而建立直线趋势预测模型。

我们知道，一次移动平均数为：

$$M_t^{(1)} = \frac{1}{N} (y_t + y_{t-1} + \cdots + y_{t-N+1})$$

在其基础上再进行一次移动平均，得到二次移动平均，其计算公式为：

$$M_t^{(2)} = \frac{1}{N} (M_t^{(1)} + \cdots + M_{t-N+1}^{(1)}) = M_{t-1}^{(2)} + \frac{1}{N} (M_t^{(1)} - M_{t-1}^{(1)})$$

当时间序列 $\{y_t\}$ 从某时期开始具有直线发展趋势，并且未来时期同样按该直线趋势变

化，我们可以设直线趋势预测模型为：

$$\hat{y}_{t+m} = a_t + b_t m, \quad m = 1, 2, \dots$$

其中， t 为当时时期数， m 为由 t 至预测期时期数； a_t 为截距， b_t 为斜率，此两者也称为平滑系数。

平滑系数可根据移动平均值来确定，其计算公式为：

$$\begin{aligned} a_t &= y_t \\ y_{t-1} &= y_t - b_t \\ y_{t-2} &= y_t - 2b_t \\ &\dots \\ y_{t-N+1} &= y_t - (N-1)b_t \end{aligned}$$

我们可以得到：

$$\begin{aligned} M_t^{(1)} &= \frac{y_t + y_{t-1} + \dots + y_{t-N+1}}{N} = \frac{y_t + (y_t - b_t) + \dots + [y_t - (N-1)b_t]}{N} \\ &= \frac{Ny_t - [1 + 2 + \dots + (N-1)]b_t}{N} = y_t - \frac{N-1}{2}b_t \end{aligned}$$

由此：

$$\begin{aligned} y_t - M_t^{(1)} &= \frac{N-1}{2}b_t \\ y_{t-1} - M_{t-1}^{(1)} &= \frac{N-1}{2}b_t \end{aligned}$$

由上式得到：

$$y_t - y_{t-1} = M_t^{(1)} - M_{t-1}^{(1)} = b_t$$

故可得：

$$M_t^{(1)} - M_t^{(2)} = \frac{N-1}{2}b_t$$

由此可得，平滑系数计算公式为：

$$a_t = 2M_t^{(1)} - M_t^{(2)}$$

$$b_t = \frac{2}{N-1} (M_t^{(1)} - M_t^{(2)})$$

将每一个像素格都作为一个时间序列样本，利用上述公式进行计算，向后滑动产生未来 10 年的土地构成图，像素值的表示与题目给的数据相同。

对于温度，本文考虑到基于大量的历史数据，我们选择采用构建 LSTM 循环神经网络模型进行预测未来两年内每月该天气事件的发生频次。在 LSTM 模型中，特定的记忆单元替换 RNN 链式结构中的隐层节点，在数据推演分析过程中具有更强的适应性，适合于分析预测天气事件未来发生情况。

LSTM 模型中的创新点在于引入信息集合的状态，如果不对状态进行修改，它就会一直向后传递。但是记忆总会有所改变，所以需要增加一些修正记忆的模块，这里则加入一些“门”的概念。“门”的作用是筛选信息，让一些信息可以通过，一些信息被门过滤无法通过，我们把信息集合的记忆单元结构进行如下可视化展示如下图所示：

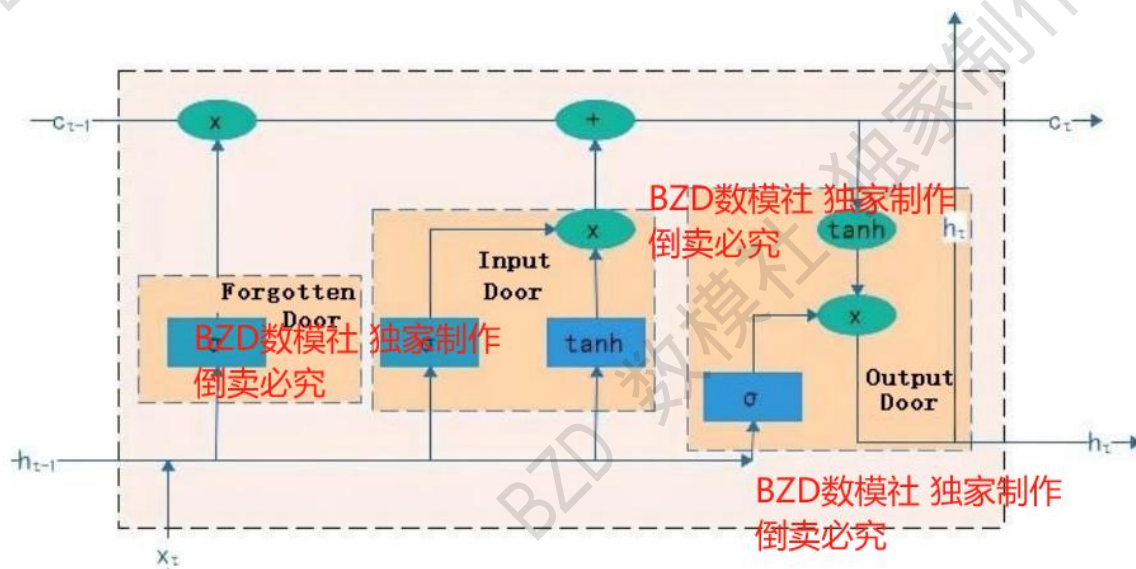


图 10 LSTM 记忆单元结构

①遗忘门

如同一个人的记忆有限，首先需要忘掉一些信息，才能记住新的信息。根据上一时刻的输出状态 h_{t-1} 和当前时刻的输入 x_t ，我们需要确定有多少长期记忆 c_{t-1} 需要去掉，遗忘的比率函数式为：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

②输入门

输入门主要负责挑选并接收新的信息。

当前时刻接收到的信息表达式为：

$$\dot{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

我们不可能一次性记住那么多信息，所以要有选择性地记住，选取的比率为：

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

因此，要放进信息集合的信息就是选择的记忆比率乘以当前时刻所接收到的信息，也就是有选择地保留一些过去的信息 $f_t \times C_{t-1}$ ，然后有选择地记住一些当前的信息 $i_t \times \dot{C}_t$ ，即：

$$C_t = f_t \times C_{t-1} + i_t \times \dot{C}_t$$

上式就是记忆的遗忘和补充的过程。

③输出门

同理，我们不能直接把信息集合中的全部信息都输出，仍然需要有选择地输出，选择率公式表达为：

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

因此，经过选择后输出的信息就是选择率乘以信息集合中的信息，表达式为：

$$h_t = o_t \times \tanh(C_t)$$

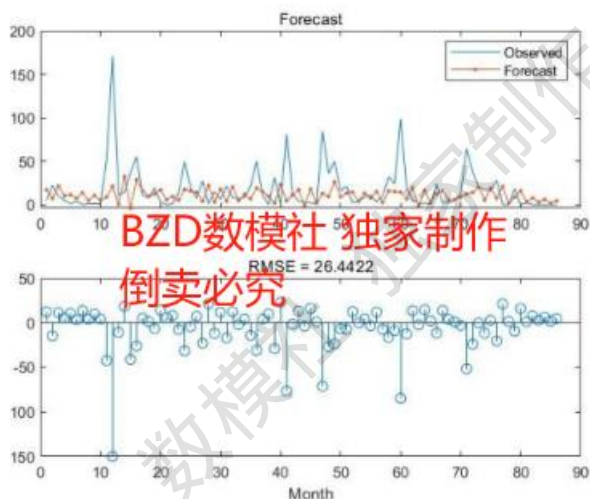
本文利用 Python 的 PyTorch 框架搭建模型，将温度数据输入模型中，仅使用前十年的数据的前 80% 的数据作为训练集，后 20% 的数据作为测试集，来评估模型的预测效果。其中每个像素点都算作一个时间序列，因此样本总个数为 144×256 ，时间序列的步数划分为 10 步，因此样本总数会扩大 365 倍，因为 10 个时间步截成一个小样本，总共十年的数

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接

<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

据，也即 365×10 个数据，以这些小样本作为小批量中的样本进行训练，这样的数据量就不会使 LSTM 出现过拟合的现象，若数据量较少，可选择参数较少的 GRU 作为备选方案，本文根据数据量确定使用 LSTM 进行预测，最终在测试集上截取的测试结果如下，仅展示 4 个像素格的预测结果，其余的不再进行画图了。



BZD数模社

2024.09.22

倒卖必究

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接

<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

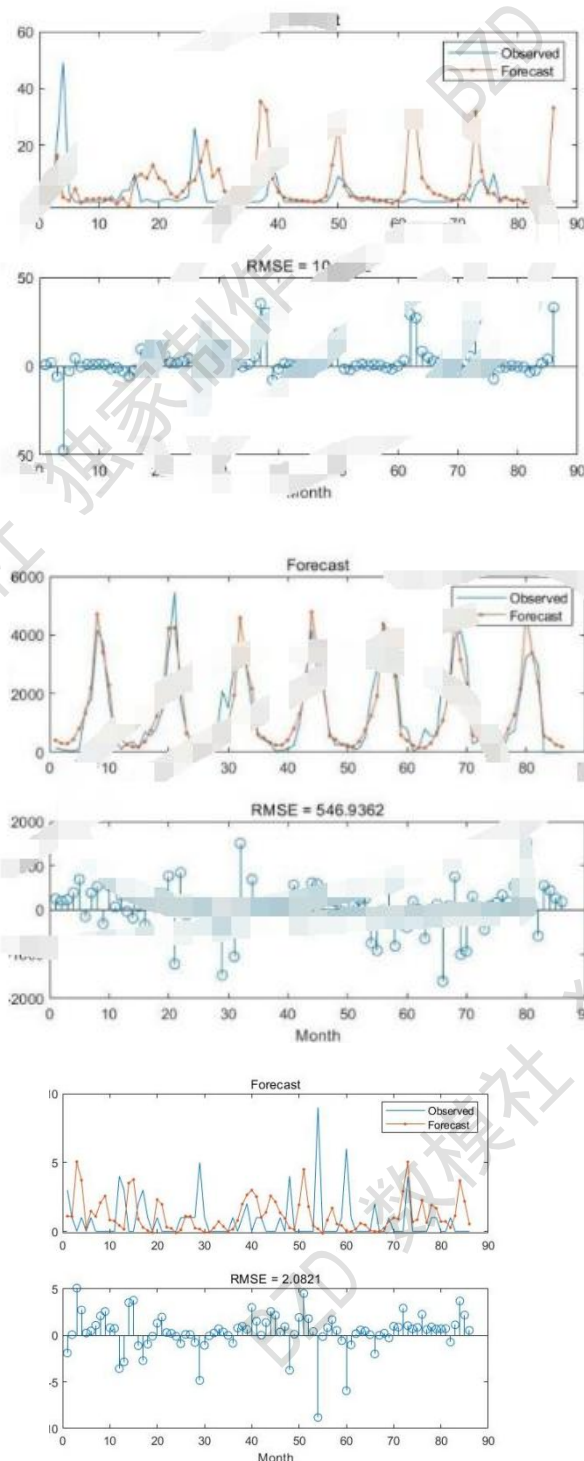


图 11 LSTM 误差评估结果图

利用预测出的数据代入前一节中建立的逻辑回归模型，预测出未来十年各个像素格发

2024.09.22

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323

倒卖必究

生降雨成灾现象的预测（由于维度较高，本文不再进行展示，仅展示最终结果），将未来 10 年出现 1 的次数超过 100 次的地区标红处理，结果如下图，其中黑色部分为国外，红色部分为未来 10 年出现 1 的次数不超过 100 次的地区，白色部分为未来 10 年出现 1 的次数超过 100 次的地区，本文仅讨论这种二分类问题，若需要推广至多分类问题，也即对全国所有城市分类出应对暴雨灾害能力最为脆弱、较为脆弱、一般、不脆弱等类的话，只需将阈值 100 调为阶梯型的阈值即可，不同研究团队可以选用不同的阈值，本文不再加入自身的主观性了。

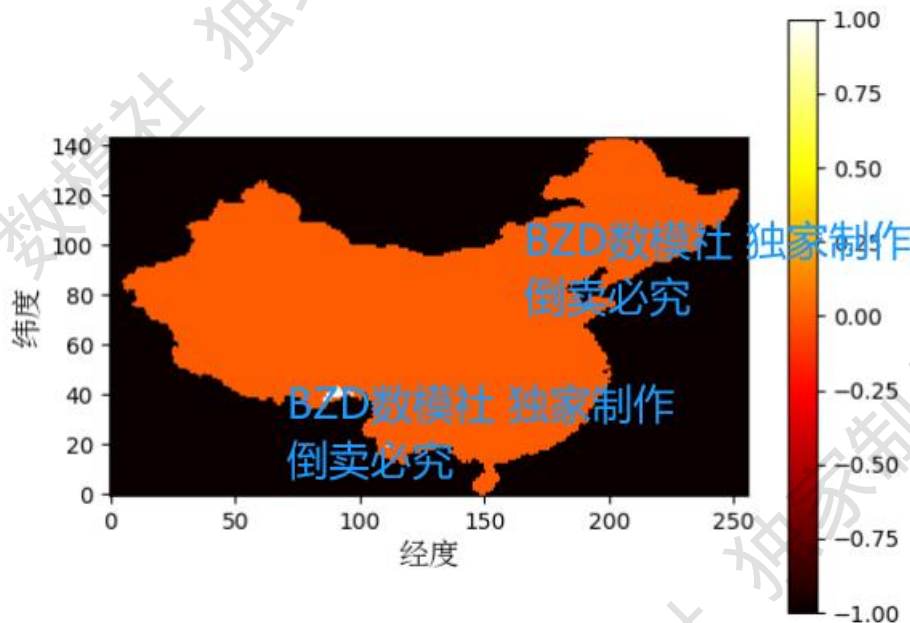


图 12 问题三结果图

因此综合上述两个模型的结果来看，（这段叙述是空缺，可以自行补充，也可以等待后续补充资料，预计 9 月 22 日晚之前，所有的结果语文叙述补充完整）

5.5 问题四模型的建立与求解

5.5.1 中国土地利用变化的影响因素确立

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群：689964173 数模交流群 2 群：275032074
国赛助攻资料通知群：928949323

BZD数模社

2024.09.22

倒卖必究

为了结合在前三问中描述、估计和预测任务中的“特性”，利用地理大数据，建立相应的数学模型，对数据进行简化和综合，描述中国土地利用变化的特征与结构，本文选择了数据集 3 的降水量、数据集 5 的人口数量、数据集 6 的 GDP 数据作为中国土地利用变化的影响因素。

自变量	降水量 X_{tij} 、人口数量 Y_{tij} 、GDP H_{tij}
因变量	土地利用变化 Z_{tij}

其中 Z_{tij} 共有五个，分别为耕地、林地、草地、灌木丛、湿地，本文不再重复分析五种植被，仅针对其中耕地作为因变量进行分析，因此问题转化为和问题一、二、三一样的问题，我们需要对其中的人口数量和 GDP 做一些描述型统计，用 1 到 3 个较为简洁的统计指标或统计图表，对这两个变量在 1990 至 2020 年间中国范围内的时空演化特征进行描述和总结。因此接下来针对以上变量重复问题一二三的工作，实现细节不再赘述，仅展示结果与结果的分析。

5.5.2 结果分析

对人口和 GDP 每五年做一个平均，画出下图得出结果，人口聚集，一线城市人多，荒郊野岭人少（这段叙述是空缺，可以自行补充，也可以等待后续补充资料，预计 9 月 22 日晚之前，所有的结果语文叙述补充完整）

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接
<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

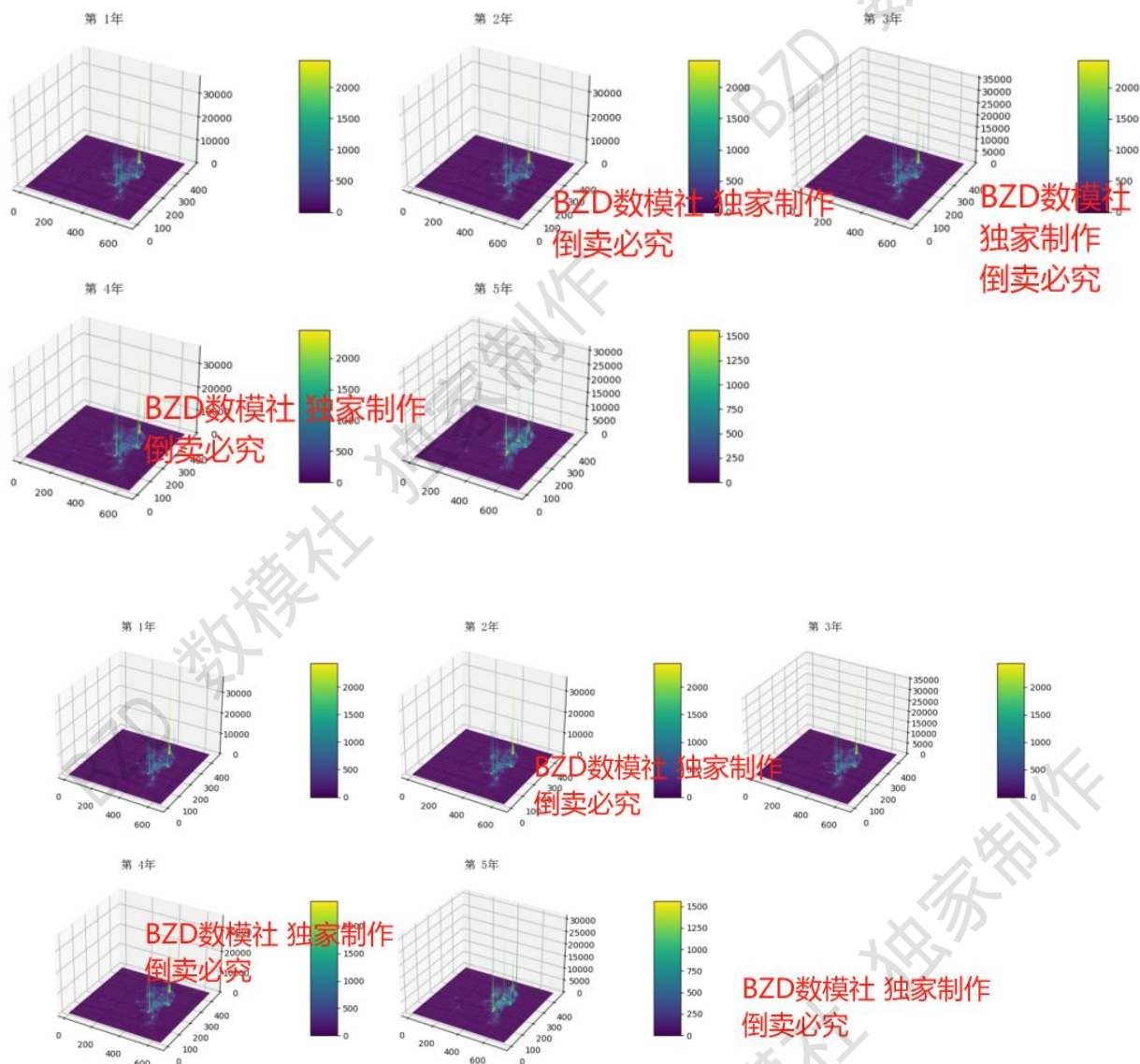


图 13 全国人口分布五年平均三维可视化

图 14 全国 GDP 分布五年平均三维可视化

数据处理方式与前文完全相同，不再赘述，逻辑回归模型如下所示：

$$P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)}}$$

其中 X_1 为降水量， X_2 为人口数量， X_3 为 GDP 数据， P 为土地的耕地覆盖率，因为覆盖率取值 0 到 1 之间，因此问题四可以沿用前文的逻辑回归模型，只不过不再取阈值变成二分类的问题，而是直接进行拟合。利用 SPSS 对逻辑回归模型中的参数进行拟合。利用 SPSS 进

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群：689964173 数模交流群 2 群：275032074
国赛助攻资料通知群：928949323

倒卖必究

行求解后，得到系数如下所示：

表 7	
参数	迭代到的局部最优解
β_0	4.7682
β_1	3.5019
β_2	-10.2503
β_3	-2.7834

（这段叙述是空缺，可以自行补充，也可以等待后续补充资料，预计 9 月 22 日晚之前，所有的结果语文叙述补充完整）

六、模型总结

6.1 模型优点

- 1、数据驱动的分析方法：本研究采用了大量地理大数据，通过数据驱动的方法来分析和预测地理环境的时空演化特征，增强了模型的实用性和可靠性。
- 2、多变量综合分析：模型综合了降雨量、地形、土地利用等多种因素，全面考虑了它们在极端天气事件中的作用，提高了预测的准确性。
- 3、时空演化特征的描述：通过构建描述性统计方法和使用统计图表，有效地描述了 1990~2020 年间中国范围内降雨量和土地利用/土地覆被类型的时空演化特征。
- 4、逻辑回归模型的应用：在多个问题中，逻辑回归模型被用来量化不同因素对极端天气事件的影响，模型结果具有较好的解释性和预测能力。
- 5、LSTM 循环神经网络的预测能力：在问题三中，使用了 LSTM 循环神经网络对温度等变量进行预测，该模型能够捕捉时间序列数据中的长期依赖关系，提高了预测的准确性。

6.2 模型缺点

- 1、数据限制：模型的准确性受到数据质量和完整性的限制，如果数据存在缺失或误差，
相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群：689964173 数模交流群 2 群：275032074
国赛助攻资料通知群：928949323

BZD 数模社

2024.09.22

倒卖必究

可能会影响模型的预测结果。

2、模型泛化能力：模型主要基于历史数据进行训练，对于未来变化的适应性和泛化能力可能有限。

3、参数选择的主观性：在逻辑回归模型中，阈值的选择具有一定的主观性，不同的阈值可能会导致不同的预测结果。

4、计算复杂性：LSTM 模型虽然预测能力强，但其计算复杂性较高，需要大量的计算资源和时间。

5、过度依赖历史数据：模型在预测未来趋势时，过度依赖历史数据，可能无法充分考虑未来可能出现的新情况或政策变化。

6.1 模型推广

1、政策制定支持：模型可以为政府在生态保护、城市规划和灾害预防等方面提供决策支持。

2、气候变化研究：模型可以应用于更广泛的气候变化研究，帮助科学家更好地理解全球气候变化对地理环境的影响。

3、资源管理优化：通过模型预测的土地利用变化特征，可以指导更有效的资源管理和分配。

4、灾害预警系统：模型可以集成到灾害预警系统中，提前预测极端天气事件，减少灾害带来的损失。

5、教育和培训：模型可以作为教育和培训工具，帮助学生和专业人员了解地理信息系统和数据分析在解决实际问题中的应用。

七、参考文献

[1] 曾悠. 大数据时代背景下的数据可视化概念研究[D]. 浙江大学, 2014.

[2] 李德仁. 展望大数据时代的地球空间信息学[J]. 测绘学报, 2016, 45(04): 379-384.

[3] 王鑫, 吴际, 刘超, 等. 基于 LSTM 循环神经网络的故障时间序列预测[J]. 北京航空航天大学学报, 2018, 44(04): 772-784. DOI:10.13700/j.bh.1001-5965.2017.0285.

- [4] 陆继翔, 张琪培, 杨志宏, 等. 基于 CNN-LSTM 混合神经网络模型的短期负荷预测方法[J]. 电力系统自动化, 2019, 43(08):131-137.
- [5] 鄂海红, 张文静, 肖思琪, 等. 深度学习实体关系抽取研究综述[J]. 软件学报, 2019, 30(06):1793-1818. DOI:10.13328/j.cnki.jos.005817.
- [6] 杨荣英, 张辉, 苗张木. 物流预测技术中的移动平均线方法[J]. 武汉理工大学学报(交通科学与工程版), 2001, (03):353-355.
- [7] 王兆军, 曾渊沧, 郝刚. 移动平均线方法的最佳步长组合的确定[J]. 高校应用数学学报 A 辑(中文版), 2000, (02):95-104. DOI:10.13299/j.cnki.amjcu.000901.
- [8] 汤国安. (2019). 中国数字高程图(1KM). 国家青藏高原数据中心. [Tang, G. (2019). Digital elevation model of China (1KM). National Tibetan Plateau / Third Pole Environment Data Center.]
- [9] Fang, S., Mao, K., Xia, X., Wang, P., Shi, J., M. Bateni, S., Xu, T., Cao, M., & Heggy, E. (2021). A Daily near-surface Air Temperature Dataset for China from 1979 - 2018 (Version 1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5502275>
- [10] Han, J., Miao, C. (2022). A new daily gridded precipitation dataset for the Chinese mainland based on gauge observations. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.21432123.v4>
- [11] 余振, Philippe Ciais, 朴世龙等. 1900-2019 年中国土地利用和覆盖变化数据集 [DS/OL]. 国家生态科学数据中心, 2022.
- [12] <https://doi.org/10.12199/nesdc.ecodb.pa.2022.11>
- [13] <https://cstr.cn/15732.11.nesdc.ecodb.pa.2022.11>
- [14] Yu, Z., Ciais, P., Piao, S., Houghton, R. A., Lu, C., Tian, H., Agathokleous, E., Kattel, G. R., Sitch, S., Goll, D., Yue, X., Walker, A., Friedlingstein, P., Jain, A. K., Liu, S., & Zhou, G. (2022). Forest expansion dominates China's land carbon sink since 1980. Nature Communications, 13(1), 5374. <https://doi.org/10.1038/s41467-022-32961-2>
- [15] 王灿, 王嘉琛. (2022). 中国历史人口空间分布公里网格数据集(1990-2015 逐年). 国家青藏高原数据中心. <https://doi.org/10.12078/2017121101> [Wang, C., Wang, J. (2022). Kilometer grid dataset of China's historical population spatial distribution (1990-2015). National Tibetan Plateau / Third Pole Environment Data Center. <https://doi.org/10.12078/2017121101>]
- [16] 徐新良. (2017). 中国人口空间分布公里网格数据集. 资源环境科学数据注册与出版系统(<http://www.resdc.cn/DOI>). DOI:10.12078/2017121101
- [17] 王灿, 王嘉琛. (2022). 中国历史 GDP 空间分布公里网格数据集(1990-2015). 国家青藏高原数据中心. <https://doi.org/10.12078/2017121102> [Wang, C., Wang, J. (2022). Kilometer grid dataset of China's historical GDP spatial distribution (1990-2015). National Tibetan Plateau / Third Pole Environment Data Center. <https://doi.org/10.12078/2017121102>]
- [18] 徐新良. (2017). 中国 GDP 空间分布公里网格数据集. 资源环境科学数据注册与出版系统(<http://www.resdc.cn/DOI>). DOI:10.12078/2017121102

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接

<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

- [19] Liu, H., Jiang, D., Yang, X., & Luo, C. (2005). Spatialization approach to 1 km grid GDP supported by remote sensing. *Geo-Inf. Sci.*, 7, 120-123.
- [20] 黄莹, 包安明, 陈曦, 刘海隆, & 杨光华. (2009). 基于绿洲土地利用的区域 GDP 公里格网化研究. *冰川冻土*, (1), 158-165.
- [21] Yi, L., Xiong, L., & Yang, X. (2006). Method of pixelizing GDP data based on the GIS. *J. Gansu Sci.*, 18, 54-58.
- [22] 陈述彭. 地理系统与地理信息系统 [J]. *地理学报*, 1991, 46(1): 1-7. <https://doi.org/10.11821/xb199101001>
- [23] 彭书时, 朴世龙, 于家烁, 刘永稳, 汪涛, 朱高峰, 董金玮, 缪驰远. 地理系统模型研究进展 [J]. *地理科学进展*, 2018, 37(1): 109-120. <https://doi.org/10.18306/dlkxjz.2018.01.012>
- [24] Peuquet, D. J. (1988). Representations of Geographic Space: Toward a Conceptual Synthesis. *Annals of the Association of American Geographers*, 78(3), 375-394. <https://doi.org/https://doi.org/10.1111/j.1467-8306.1988.tb00214.x>
- [25] 胡焕庸. 中国人口之分布——附统计表与密度图 [J]. *地理学报*, 1935, 2(2): 33-74 <https://doi.org/10.11821/xb193502002>
- [26] Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3), 255-261. <https://doi.org/10.1177/2043820613513121>
- [27] 阎国年, 周成虎, 林琿, 陈旻, 乐松山, 温永宁. 地理综合研究方法的发展与思考 [J]. *科学通报*, 2021, 66(20): 2542-2554. <https://doi.org/10.1360/TB-2020-0799>

BZD数模社

2024.09.22

倒卖必究

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323

八、附录

附录 1

介绍：论文中提到的代码

```
import xarray as xr
import matplotlib.pyplot as plt
import numpy as np
from matplotlib.font_manager import FontProperties

data_path = "E:\yansai_data\CHM_PRE_0.25dg_19612022.nc"
ds = xr.open_dataset(data_path)

my_longitude = ds["longitude"].to_numpy()
my_latitude = ds["latitude"].to_numpy()
my_years = ds["years"].to_numpy()
my_time = ds["time"].to_numpy()
my_pre = ds["pre"].to_numpy()

my_pre = np.clip(my_pre, 0, None)
sys_data = my_pre

start_index = 11322
step = 365

# 设置后两个维度的区间，计算所有纬度和经度的平均值
# 1 使用全部数据，设置为 (0, 144) 和 (0, 256)
# lat_range = (0, 144)
# lon_range = (0, 256)

## shandong
# lat_range = (65, 85)
# lon_range = (170, 203)

## beijing
# lat_range = (85, 93)
# lon_range = (173, 193)

## xizang
# lat_range = (70, 125)
```

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接
<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

```
# lon_range = (125, 160)

## jilin
# lat_range = (93, 119)
# lon_range = (203, 240)

## jiangsu
# lat_range = (45, 65)
# lon_range = (175, 200)

# sichuan
lat_range = (35, 65)
lon_range = (125, 160)

# 用于存储计算结果的列表
average_values = []

# 循环遍历第一个维度
for i in range(start_index, len(sys_data), step):
    # 切片获取特定时间段的数据
    segment = sys_data[i:i + step] # 从 i 开始，取 step 个数据
    avg = segment[:, lat_range[0]:lat_range[1] + 1, lon_range[0]:lon_range[1] + 1].mean()
    average_values.append(avg)

print(average_values)
average_values = average_values[0:len(average_values)-1]
# 创建一个时间序列，假设每个数据块代表一年，从起始年份开始
start_month = 1990 # 假设数据的开始年份
num_months = len(average_values) # 根据平均值的数量确定年份数量
years = [start_month + i for i in range(num_months)]

font = FontProperties(fname=r"c:\windows\fonts\simsum.ttc", size=14)
# 绘制折线图
plt.figure(figsize=(10, 5)) # 设置图形的大小
plt.plot(years, average_values, marker='o')

# 添加标题和标签
plt.title('四川年降水量', fontproperties=font)
plt.xlabel('年份', fontproperties=font)
plt.ylabel('年平均降水量', fontproperties=font)
# 显示网格
```

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323

BZD 数模社

2024.09.22

倒卖必究

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接

<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

```
plt.grid(True)

## 显示图例
# plt.legend(['Average Value'])

# 定义保存路径
save_path = r'E:\yansai_data\month_all_sichuan.png' # 确保这个路径是存在的，或者你的程序有
权限创建它
plt.savefig(save_path)
# 显示图形
plt.show()
# print(my_longitude.shape)
# print(my_latitude.shape)
# print(my_years.shape)
# print(my_time.shape)
# print(my_pre.shape)
# variable_names = list(ds.variables.keys())
# print(variable_names)
### 查看数据集的维度
## print(ds.dims)
##
## print("-----")
### 查看数据集中的变量
## print(ds.variables)
##
## print(type(ds))

import xarray as xr
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
from matplotlib.font_manager import FontProperties

data_path = "E:\yansai_data\CHM_PRE_0.25dg_19612022.nc"
ds = xr.open_dataset(data_path)

my_longitude = ds["longitude"].to_numpy()
my_latitude = ds["latitude"].to_numpy()
my_years = ds["years"].to_numpy()
my_time = ds["time"].to_numpy()
```

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323

BZD 数模社

2024.09.22

倒卖必究

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接
<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

```
my_pre = ds["pre"].to_numpy()

my_pre = np.clip(my_pre, 0, None)

font = FontProperties(fname=r"c:\windows\fonts\simsum.ttc", size=14)
data = my_pre[20000]

# 生成 x 和 y 坐标网格
x = np.arange(data.shape[1])
y = np.arange(data.shape[0])
x, y = np.meshgrid(x, y)

# 创建一个图形和一个轴对象
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

# 绘制三维表面图
surf = ax.plot_surface(x, y, data, cmap='viridis')

# 添加颜色条
fig.colorbar(surf)

# 添加标题和轴标签
ax.set_title('某天的降水量', fontproperties=font)
ax.set_xlabel('经度', fontproperties=font)
ax.set_ylabel('纬度', fontproperties=font)

save_path = r'E:\yansai_data\annual_average_values222.png' # 确保这个路径是存在的, 或者你的
程序有权限创建它
plt.savefig(save_path)

# 显示图形
plt.show()

import xarray as xr
import matplotlib.pyplot as plt
import numpy as np
from matplotlib.font_manager import FontProperties
```

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323

BZD 数模社

2024.09.22

倒卖必究

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接
<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

```
data_path = "E:\\yansai_data\\CHM_PRE_0.25dg_19612022.nc"
ds = xr.open_dataset(data_path)

my_longitude = ds["longitude"].to_numpy()
my_latitude = ds["latitude"].to_numpy()
my_years = ds["years"].to_numpy()
my_time = ds["time"].to_numpy()
my_pre = ds["pre"].to_numpy()

my_pre = np.clip(my_pre, 0, None)
start = my_pre.shape[0] - 365*20
font = FontProperties(fname=r"c:\windows\fonts\simsun.ttc", size=14)
sum1 = my_pre[start:start+365*10].sum(axis=0)
sum2 = my_pre[start+365*10:start+365*20].sum(axis=0)

plt.imshow(sum2, cmap='hot', interpolation='nearest')
plt.colorbar() # 显示颜色条
plt.title('十年的年平均降水量', fontproperties=font)
plt.xlabel('经度', fontproperties=font)
plt.ylabel('纬度', fontproperties=font)

save_path = r'E:\\yansai_data\\kk2.png' # 确保这个路径是存在的，或者你的程序有权限创建它
plt.savefig(save_path)
plt.show()

import numpy as np
import matplotlib.pyplot as plt
from PIL import Image
from matplotlib.font_manager import FontProperties

l = []
for i in range(1900, 2020):
    path = "E:\\yansai_data\\数据实体\\forest-" + str(i) + ".tif"
    img = Image.open(path)
    img_array = np.array(img)
    l.append(img_array)

cropland_data = np.array(l)
```

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323

BZD数模社

2024.09.22

倒卖必究

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接

<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

```
data_list = []
for i in range(12):
    data_list.append(cropland_data[i:10*(i+1)].mean(axis=0))

# 假设你有 12 个二维数组，这里随机生成作为示例
print(len(data_list))
print(data_list[0].shape)

# 创建一个 3x4 的子图布局
fig, axes = plt.subplots(nrows=3, ncols=4, figsize=(15, 10)) # 可以调整 figsize 来改变整个图的大小

font = FontProperties(fname=r"c:\windows\fonts\simsum.ttc", size=14)
# 遍历每个子图和数据集，绘制热力图
for i, ax in enumerate(axes.flat):
    # 绘制热力图
    im = ax.imshow(data_list[i], cmap='hot', interpolation='nearest', origin='lower')
    fig.colorbar(im, ax=ax, pad=0.02) # 为每个子图添加颜色条
    ax.set_title(f'第 {i+1} 个十年', fontproperties=font) # 可以为每个子图设置标题

# 调整子图之间的间距
plt.tight_layout()

# 显示整个图
plt.show()

import xarray as xr
import matplotlib.pyplot as plt
import numpy as np
from matplotlib.font_manager import FontProperties

data_path = "E:\yansai_data\CHM_PRE_0.25dg_19612022.nc"
ds = xr.open_dataset(data_path)

my_longitude = ds["longitude"].to_numpy()
my_latitude = ds["latitude"].to_numpy()
my_years = ds["years"].to_numpy()
my_time = ds["time"].to_numpy()
my_pre = ds["pre"].to_numpy()
```

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323

BZD 数模社

2024.09.22

倒卖必究

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接
<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

```
# my_pre = np.clip(my_pre, 0, None)
start = my_pre.shape[0] - 365*20
font = FontProperties(fname=r"c:\windows\fonts\simsum.ttc", size=14)
sum1 = my_pre[start:start+365*10].mean(axis=0)

for i in range(sum1.shape[0]):
    for j in range(sum1.shape[1]):
        if sum1[i,j] < 7.5 and sum1[i,j] >= 0:
            sum1[i,j] = 0
        if sum1[i,j] > 7.5:
            sum1[i,j] = 1
        if sum1[i,j] < 0:
            sum1[i,j] = -1
max_values = np.max(sum1, axis=1)
min_values = np.min(sum1, axis=1)
print(max_values)
print(min_values)

# sum2 = my_pre[start+365*10:start+365*20].sum(axis=0)

plt.imshow(sum1, cmap='hot', interpolation='nearest', origin='lower')
plt.colorbar() # 显示颜色条
plt.xlabel('经度', fontproperties=font)
plt.ylabel('纬度', fontproperties=font)

save_path = r'E:\yansai_data\kk3333.png' # 确保这个路径是存在的，或者你的程序有权限创建它
plt.savefig(save_path)
plt.show()

import numpy as np
import matplotlib.pyplot as plt
from PIL import Image
from matplotlib.font_manager import FontProperties
font = FontProperties(fname=r"c:\windows\fonts\simsum.ttc", size=14)
data_list = []
for i in range(2011, 2016):
    path = ("E:\\yansai_data\\pop_output\\pop"+str(i)+"\\pop"+str(i)+".tif")
    img = Image.open(path)
```

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323

BZD 数模社

2024.09.22

倒卖必究

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接

<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

```
img_array = np.array(img)
img_array = np.clip(img_array, 0, None)
img_array = img_array[:, :10, :10]
data_list.append(img_array)

# 创建一个 5*5 的子图布局
fig = plt.figure(figsize=(15, 10)) # 可以调整 figsize 来改变整个图的大小
nrows = 2
ncols = 3

for i, data in enumerate(data_list):
    # 绘制热力图
    ax = fig.add_subplot(nrows, ncols, i + 1, projection='3d')
    # 生成 x 和 y 坐标网格
    x = np.arange(data.shape[1])
    y = np.arange(data.shape[0])
    x, y = np.meshgrid(x, y)

    # 绘制三维表面图
    surf = ax.plot_surface(x, y, data, cmap='viridis')
    # 添加颜色条
    fig.colorbar(surf, ax=ax, shrink=0.5, aspect=5, pad=0.25) # 调整颜色条的大小和比例

    # 添加标题和轴标签
    ax.set_title(f'第 {i+1} 年', fontproperties=font) # 可以为每个子图设置标题

# 调整子图之间的间距
plt.tight_layout()
# save_path = r'E:\yansai_data\cc33.png' # 确保这个路径是存在的，或者你的程序有权限创建它
# plt.savefig(save_path)
# 显示整个图
plt.show()

# 创建一个二维数组，例如一个随机数组
# data = my_pre[20000] # 10x10 的随机数组

## 生成 x 和 y 坐标网格
```

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群：689964173 数模交流群 2 群：275032074
国赛助攻资料通知群：928949323

BZD 数模社

2024.09.22

倒卖必究

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接
<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

```
# x = np.arange(data.shape[1])
# y = np.arange(data.shape[0])
# x, y = np.meshgrid(x, y)
#
## 创建一个图形和一个轴对象
# fig = plt.figure()
# ax = fig.add_subplot(111, projection='3d')
#
## 绘制三维表面图
# surf = ax.plot_surface(x, y, data, cmap='viridis')
#
## 添加颜色条
# fig.colorbar(surf)
#
## 添加标题和轴标签
# ax.set_title('某天的降水量')
# ax.set_xlabel('经度')
# ax.set_ylabel('纬度')
# ax.set_zlabel('降水量')
#
## 显示图形
# plt.show()

import numpy as np
import matplotlib.pyplot as plt
from PIL import Image
from matplotlib.font_manager import FontProperties
font = FontProperties(fname=r"c:\windows\fonts\simsun.ttc", size=14)
data_list = []
for i in range(2010, 2015):
    path = ("E:\\yansai_data\\GDP_output\\gdp"+str(i)+".tif")
    img = Image.open(path)
    img_array = np.array(img)
    img_array = np.clip(img_array, 0, None)
    img_array = img_array[:, :10, :10]
    data_list.append(img_array)

# 创建一个 5*5 的子图布局
```

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323

BZD 数模社

2024.09.22

倒卖必究

```
fig = plt.figure(figsize=(15, 10)) # 可以调整 figsize 来改变整个图的大小
nrows = 2
ncols = 3

for i, data in enumerate(data_list):
    # 绘制热力图
    ax = fig.add_subplot(nrows, ncols, i + 1, projection='3d')
    # 生成 x 和 y 坐标网格
    x = np.arange(data.shape[1])
    y = np.arange(data.shape[0])
    x, y = np.meshgrid(x, y)

    # 绘制三维表面图
    surf = ax.plot_surface(x, y, data, cmap='viridis')
    # 添加颜色条
    fig.colorbar(surf, ax=ax, shrink=0.5, aspect=5, pad=0.25) # 调整颜色条的大小和比例

    # 添加标题和轴标签
    ax.set_title(f'第 {i+1} 年', fontproperties=font) # 可以为每个子图设置标题

# 调整子图之间的间距
plt.tight_layout()
# save_path = r'E:\yansai_data\cc33.png' # 确保这个路径是存在的，或者你的程序有权限创建它
# plt.savefig(save_path)
# 显示整个图
plt.show()
```

附录 2

介绍：支撑材料的文件列表

数据预处理.py
数据集 1.py
数据集 2.py
数据集 3.py
数据集 4.py
数据集 5.py
数据集 6.py

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群：689964173 数模交流群 2 群：275032074
国赛助攻资料通知群：928949323

BZD 数模社

2024.09.22

倒卖必究

论文写作过程中遗留数个致命问题，无关代码，该问题解决方式仅在官网授权售后群答疑，盗卖方式购买资料不提供答疑，感谢理解

【腾讯文档】2024 年华为杯-研赛资料助攻说明约定链接
<https://docs.qq.com/doc/p/e03d6bffc00b392932b7dc1fbcf746ef9b086a8a>

--

BZD数模社

2024.09.22

倒卖必究

相关资料 可以加群 或 关注 公众号 BZD 数模社
数模交流群 1 群: 689964173 数模交流群 2 群: 275032074
国赛助攻资料通知群: 928949323