



西北农林科技大学

# 博士学位论文

气候变化对作物生长和产量影响的模拟及不确定性研究

学科专业 土地资源与空间信息技术

研究方向 气候变化与农业

论文作者 李林超

指导教师 于强

联合指导教师 \_\_\_\_\_

论文提交时间 2023年5月





Dissertation Submitted to Northwest A & F University  
in Partial Fulfillment of the Requirements  
for Degree of  
Doctor of Philosophy

# Crop Growth and Yield Predictions and Uncertainty Analysis under Climate Change

Major: Agricultural Resources and Environment

Research Field: Land Resources and Spatial Information Technology

Candidate: Linchao Li

Supervisor: Qiang Yu

Co- Supervisor:

Date of submission: May 2022

College of Natural Resources and Environment  
Northwest A & F University



分类号：S127

UDC：631

密级：公开

学校代码：10712

研究生学号：2019060331

## 西北农林科技大学博士学位论文

# 气候变化对作物生长和产量影响的模拟及不确定性研究

论文作者：李林超

指导教师：于强

答辩委员会：

西北农林科技大学水利与建筑工程学院 李 毅 研究员（主席）

西北农林科技大学水利与建筑工程学院 何建强 教授

西北农林科技大学资源环境学院 齐雁冰 教授

中国农业大学资源与环境学院 王 靖 教授

中国农业大学土地科学与技术学院 冯璞玉 副教授

答辩日期：2023年5月17日

本研究得到国家自然科学基金委员会与美国国家科学基金会“食品、能源、水”系统关联合作研究项目：全球变化背景下“粮食-能源-水”耦合系统的多尺度模拟与整体优化：黄河与密西西比河流域可持续性比较研究（编号：41961124006）资助。



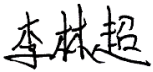
## 研究生学位论文的独创性声明

本人声明：所呈交的学位论文是我个人在导师指导下独立进行的研究工作及取得的研究结果；论文中的研究数据及结果的获得完全符合学校《关于规范西北农林科技大学研究生学术道德的暂行规定》，如果违反此规定，一切后果与法律责任均由本人承担。

尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究结果，也不包含其他人和自己本人已获得西北农林科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文的致谢中作了明确的说明并表示了谢意。

研究生签名：  时间： 2023 年 5 月 19 日

## 导师指导研究生学位论文的承诺

本人承诺：我的研究生  所呈交的学位论文是在我指导下独立开展研究工作及取得的研究结果，属于我现岗职务工作的结果，并严格按照学校《关于规范西北农林科技大学研究生学术道德的暂行规定》而获得的研究结果。如果违反学校《关于规范西北农林科技大学研究生学术道德的暂行规定》，我愿接受按学校有关规定的处罚处理并承担相应导师连带责任。

导师签名：  时间： 2023 年 5 月 23 日



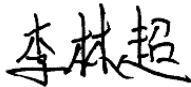
## 关于研究生学位论文使用授权的说明


本学位论文的知识产权归属西北农林科技大学。本人同意西北农林科技大学保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅；同意西北农林科技大学将本学位论文的全部或部分授权汇编录入《中国博士/硕士学位论文全文数据库》和《中国学位论文全文数据库》进行出版，并享受相关权益。

本人保证，在毕业离开（或者工作调离）西北农林科技大学后，发表或者使用本学位论文及其相关的工作成果时，将以西北农林科技大学为第一署名单位，否则，愿意按《中华人民共和国著作权法》等有关规定接受处理并承担法律责任。

任何收存和保管本论文各种版本的其他单位和个人(包括研究生本人)未经本论文作者的导师同意，不得有对本论文进行复制、修改、发行、出租、改编等侵犯著作权的行为，否则，按违背《中华人民共和国著作权法》等有关规定处理并追究法律责任。

（保密的学位论文在保密期限内，不得以任何方式发表、借阅、复印、缩印或扫描复制手段保存、汇编论文）

研究生签名：  时间：2023年5月23日

导师签名：  时间：2023年5月23日







## 摘要

量化气候因素对作物生产力的影响是一个国内外普遍关注的重大科学问题。近年来, 极端气候事件(例如暴雨、干旱和高温等) 趋多趋强, 气候风险水平呈上升趋势。因此, 对产量变化的及时准确预测和量化气候变化对产量的影响的评估对于制定适应措施以及全球粮食安全具有重要意义。基于过程的作物模型是用来解释当地环境变量、作物基因型和管理实践之间相互作用的一种有效工具。由于作物模型对某些过程过于简化, 以及存在一些参数的不确定性, 可能会导致预测结果的不准确。作物模型往往需要大量的当地观测数据来校准模型, 且需要较大的计算资源。基于统计模型的机器学习算法易于处理, 且相对容易计算。可以较好的捕捉到环境变量与产量的非线性关系, 具有较高的准确性。然而机器学习方法不能充分的考虑作物生长过程(例如  $\text{CO}_2$  施肥效应), 在分析未来气候变化对产量影响方面具有较大的局限性。因此, 需要考虑采用不同的方法(过程模型与统计模型) 来更加全面、准确地解释气候变化与作物生产力之间的关系, 并且合理量化分析的不确定性。这将有助于制定更加科学、有效的适应措施, 以确保全球粮食安全。

本文采用不同方法在区域、全国以及全球尺度上进行产量变化和气候变化对产量的影响的模拟和预测。首先, 通过机器学习对产量进行预测和模拟, 并确定不同变量对产量的影响并在此基础上构建产量预报模型以更早的在收获期前估算产量。然后基于作物模型预测未来气候变化对产量的影响并量化不确定性来源。在此基础上, 将两种方法结合, 以提高模型的精度并有效地约束模型的不确定性。最后, 将机器学习结合作物模型的方法进行扩展, 建立了大尺度环流与产量的遥相关关系, 并分析了气候驱动因子对产量的影响。本研究可以为决策者制定科学的管理措施提供有价值的参考, 并有助于提高对气候变化与粮食安全之间关系的认识。本论文的主要结论如下:

(1) **基于机器学习预测作物产量及其与环境变量的响应关系。**通过建立基于机器学习方法的小麦产量估算模型, 比较了不同机器学习方法和数据集的预测能力。结果表明随机森林模型在小麦产量模拟方面性能优于支持向量机。使用基于植被的指数  $\text{NIR}_v$  ( $R^2 = 0.74$ ;  $\text{RMSE} = 758 \text{ kg/ha}$ ) 的随机森林模型比使用基于  $\text{EVI}$  ( $R^2 = 0.73$ ;  $\text{RMSE} = 762 \text{ kg/ha}$ ) 或  $\text{NDVI}$  ( $R^2 = 0.73$ ;  $\text{RMSE} = 770 \text{ kg/ha}$ ) 的模型更为精确。本研究进一步确定了各种环境变量对小麦产量的影响程度。结果表明, 基于植被的指数对小麦产量的影响最大, 其中 T3 期间(拔节-抽穗)的  $\text{NIR}_v$  是最重要的预测因子。此外, 辐射(时段: 分蘖-抽穗)、相对湿度(时段: 播种-分蘖)、土壤有机碳(SOC)、风速(时段: 整个生育阶段)、小雨(时段: 播种-拔节)和霜冻(时段: 分蘖-拔节)

被确定为限制小麦产量的主要因素。研究还发现,在本研究中,灌溉减轻了干旱对产量变化的影响。本研究采用局部依赖曲线解释了每个发育时期的不同预测因子如何影响小麦产量,发现 NIRv 与小麦产量既有线性关系也有非线性关系,而小麦产量对其他环境变量则呈现类似阈值的反应。这些局部依赖关系图的结果有助于更好地了解这些因素如何限制小麦产量。最后,本研究证明了使用 NIRv 进行产量预测可以得到相对较好的精度,并且所提出的建模方法使用公开数据广泛适用于全球其他地区。

**(2) 建立动态产量早期预报模型并识别关键生长阶段。**利用机器学习方法,通过多源环境变量和中国不同的田间试验点,构建了提前预报产量的产量预报模型。研究表明,使用由多源环境变量驱动的机器学习可以提供精度较高的作物产量预测。研究发现在收获前大约一到三个月的产量可以得到较好的模拟精度,冬小麦 ( $r = 0.81-0.85$ ,  $nRMSE = 10.5-11.4\%$ ); 春玉米 ( $r = 0.79-0.81$ ,  $nRMSE = 17.1-17.9\%$ )、夏玉米 ( $r = 0.77-0.79$ ,  $nRMSE = 10.2-10.4\%$ )、早稻 ( $r = 0.71-0.72$ ,  $nRMSE = 7.4-7.5\%$ ), 中稻 ( $r = 0.78-0.82$ ,  $nRMSE = 7.6-8.3\%$ ), 晚稻 ( $r = 0.76-0.78$ ,  $nRMSE = 8.6-8.9\%$ )。此外,分析了决定小麦、玉米和水稻产量的主要预测因子。一般来说,太阳辐射和植被指数(特别是在拔节到乳熟期)被确定为冬小麦的主要预测因子;植被指数(整个生长季节)和干旱(尤其是在出苗到抽穗阶段)是春玉米最重要的预测指标;土壤水分(整个生长季节)是夏玉米、晚稻和中稻的主要预测因子;降水(特别是在孕穗期到抽穗期)是早稻的主要预报因子。未来的工作将开发一种混合方法,使用生物物理模型和机器学习技术来提高作物产量预测的准确性。

**(3) 基于多模型集合预测气候变化对作物产量的影响及不确定性分析。**基于不同的网格作物模型(GGCM)和GCM预测未来气候变化对产量的影响。同时分析了作物产量变化方面的集合配置(模型组成和大小)对建模不确定性的影响。结果显示,气候变化下建模不确定性的主要来源在不同GCM和GGCM组合的集合中存在较大的差异,尤其是对于玉米、水稻和大豆等作物,而小麦的不确定性相对较小。此外,随机选择大约6个GGCM和10个GCM可以确定9个作物模型和32个气候模型的建模不确定性。然而,基于聚类分析的结果可以进一步减少最小集合数目,发现在保证每一个聚类至少有一个模型的情况下,大约3-4个模型即可反映整体的方差。这些结果突出了模型组成和集合大小在确定作物产量预测中不确定性的主要来源方面的重要性。

**(4) 基于机器学习提高网格作物模型对产量的预测并约束不确定性。**结合网格化作物模型和机器学习,考虑了不同因素(如极端气候和病虫害风险)构建混合模型,并分析了历史和未来不同排放情景下气候变化对玉米和大豆产量的影响。研究发现通过结合机器学习来校准作物模型输出,可以显著提高模型的精度。在校准前,玉米的 $R^2$ 为0.06-0.37, $nRMSE$ 大约为0.2-0.4(仅有JULES大于50%);大豆的 $R^2$ 为0.12-0.48,

nRMSE 大约为 0.2-0.35。而对于混合模型，玉米的  $R^2$  为 0.37-0.58，nRMSE 全部小于 0.2；大豆的  $R^2$  为 0.31-0.47，nRMSE 基本低于 0.2。病虫害、高温和干旱是大豆产量的主要影响因素；而低温、病虫害和干旱则是玉米产量变化的主要影响因素。本研究的方法可以约束 25.8-74.1%的不确定性。在 T1 时段（2040-2069），作物模型是玉米产量不确定性的主要来源，而气候模式则是大豆产量不确定性的主要来源。在 T2 时段（2070-2099），排放情景成为主要的不确定性来源。值得注意的是，不确定性来源存在一定的区域差异。本研究改进了网格模型的预测能力，并成功约束了作物模型的不确定性，同时揭示了不确定性来源的区域差异。

**（5）未来情景下气候驱动因子对作物产量的影响。**将统计模型结合作物模型的方法进行拓展，以进一步探究气候驱动因素对作物产量的影响。结果表明，在历史期间，厄尔尼诺与南方涛动（ENSO）是小麦和玉米产量的主要驱动因素，对欧洲和亚洲北部地区影响最大；而印度洋偶极子（IOD）则是非洲和南美洲地区产量的主要驱动因素。未来期间（2000-2099），NAO 在北半球的玉米产量中占比例不断增大，而在水稻和大豆中，北大西洋涛动（NAO）同样有着明显的提升。在 SSP585 情景下，IOD、ENSO 和 NAO 的正向异常导致产量有非常强烈的变化，而负向异常变化不大，尤其是 IOD 和 NAO。在 SSP126 阶段，产量变化与历史相似，但 ENSO 异常仍然导致产量有明显的变化。对于玉米来说，NAO 的负向异常和 ENSO 正向和负向的异常将导致多个区域的产量受到威胁。小麦则应该特别关注 IOD 异常。对于大豆和水稻，NAO 的正向或负向异常将对产量造成极大的减产。在不确定性方面，GCM 的不确定性对产量变化的贡献大于 SSP，而 SSP 的不确定性贡献大于 GGCM。除了在水稻和玉米产量对 NAO 的敏感性分析中，SSP 的不确定性贡献高于 GCM。

**关键词：**机器学习；产量预测；气候变化；气候模式；作物模型；不确定性分析



## ABSTRACT

The impact of quantified climate factors on crop productivity is a significant scientific issue that has attracted widespread attention both domestically and internationally. In recent years, extreme weather events such as heavy rain, drought, and high temperatures have become more frequent and intense, and the level of climate risk is on the rise. Therefore, timely and accurate predictions of yield changes and evaluations of the impact of quantified climate changes on yield are of great importance for developing adaptation measures and ensuring global food security. Process-based crop models are effective tools for explaining the interactions between local environmental variables, crop genotypes, and management practices. However, the simplification of certain processes and uncertainty surrounding certain parameters in these models may result in inaccurate predictions. Furthermore, crop models often require a large amount of local observation data for model calibration and significant computational resources. Machine learning algorithms based on statistical models are easy to handle and relatively easy to compute. They can capture the nonlinear relationship between environmental variables and yield with high accuracy. However, these algorithms cannot fully consider the crop growth process (such as the CO<sub>2</sub> fertilization effect) and have significant limitations in analyzing the impact of future climate change on yield. Therefore, we need to consider using different methods to comprehensively and accurately explain the relationship between climate change and crop productivity, and reasonably quantify the uncertainty of the analysis. This will help us develop more scientific and effective adaptation measures to ensure global food security.

In this study, we use different methods to simulate and predict yield changes and the impact of climate change on yield at different scales. Firstly, we use machine learning to predict and simulate yield and determine the impact of different variables on yield. Then, we combine machine learning with grid crop models to improve the accuracy of the model and effectively constrain the model's uncertainty. Based on this, we propose a new framework to explore yield prediction and uncertainty and study the impact of model combination and the number of models on uncertainty. Finally, we establish a remote correlation between large-scale circulation and yield and analyze the impact of climate variability on food security. This study provides valuable reference for decision-makers to develop scientific management measures and helps to improve understanding of the relationship between climate change and food security. The main conclusions of this paper

are as follows:

**(1) Developing machine learning models to predict wheat yield and reveal their non-linear response.** We have developed machine learning-based models to accurately predict wheat yield by incorporating multi-source environmental data, such as soil properties, climate, and vegetation indices. Our research found that the Random Forest (RF) model outperformed the Support Vector Machine (SVM) model in predicting wheat yield. Additionally, we discovered that the RF model using Near-Infrared Vegetation Index (NIRv) had a slightly better prediction accuracy ( $R^2=0.74$ ; RMSE=758 kg/ha) than using Enhanced Vegetation Index (EVI) ( $R^2=0.73$ ; RMSE=762 kg/ha) or Normalized Difference Vegetation Index (NDVI) ( $R^2=0.73$ ; RMSE=770 kg/ha). Our findings also showed that vegetation-based indices had the most significant impact on wheat yields compared to other environmental covariates. Based on our RF\_M5 model, we identified that NIRv during T3 was the most critical predictor for determining crop yield. Additionally, we identified that several environmental factors such as Rad (T2 and T3), RHum (T1), SOC, Ws (T1–T4), R5 (T1–T2), and Frost (T2) were the main limiting factors for wheat yield. Interestingly, drought had a relatively small contribution to yield change due to the use of irrigation during winter wheat production in our study. We constructed Partial Dependence Plot (PDP) plots to explain how different predictors at each developmental period affected wheat yield based on RF models. We discovered that NIRv showed both linear and nonlinear relationships with wheat yield, while wheat yield had threshold-like responses to other environmental variables. These PDP results can help better understand how factors limit wheat yield. Our findings demonstrate the potential of using NIRv for yield prediction, and our modelling approach can be applied globally in other regions using publicly available data. In the future, incorporating process-based crop models or newly developed vegetation indices could further enhance our yield prediction model.

**(2) Crop yield forecasting and associated optimum lead time analysis based on multi-source environmental data.** We have developed yield forecasting systems for three major crops using a machine learning method that using multi-source environmental variables and field trial data from diverse locations across China. Our findings indicate that our machine learning approach driven by multi-source environmental data can provide satisfactory crop yield forecasts, with performance comparable to that reported in previous studies. For winter wheat, we were able to predict yields with a correlation coefficient ( $r$ ) of 0.81-0.85 and normalized root-mean-square error (nRMSE) of 10.5–11.4% around one to three months prior to harvest. For spring maize, summer maize, early rice, mid rice, and late rice, we were able to predict yields with  $r$  ranging from 0.71-0.82 and nRMSE ranging



from 7.4-17.9%, one to two months before harvest. Our machine learning model provides useful information for farmers and policymakers to reduce yield loss before the end of the growing season. Furthermore, we identified the key predictors that influence the yields of wheat, maize, and rice. For winter wheat, we found that solar radiation and vegetation indices (especially during the jointing to milk development stages) were the main predictors, while for spring maize, vegetation indices (throughout the growing season) and drought (especially during the emergence to tasseling stages) were the most important predictors. Soil moisture (throughout the growing season) was the dominant predictor for summer maize, late rice, and mid rice, while precipitation (especially during booting to heading stages) was the main predictor for early rice. Our future work will involve the development of a hybrid approach that combines a biophysical model with our machine learning technique to further improve the accuracy of crop yield forecasts.

**(3) Project crop yield changes and uncertainty analysis under future climate change using multi-model ensemble approaches.** Linked climate and crop simulation models are widely used to assess the impact of climate change on agriculture. However, it is unclear how ensemble configurations (model composition and size) influence crop yield projections and uncertainty. Here, we investigate the influences of ensemble configurations on crop yield projections and modelling uncertainty from Global Gridded Crop Models (GGCMs) and Global Climate Models (GCMs) under future climate change. We found that the projections of crop yields and their related uncertainty are subject to variations depending on different ensemble configurations and exhibit regional specificity, particularly for crops such as wheat and soybean that are sensitive to ensemble members. Ignoring this difference could result in an underestimation of the impact of climate change on crop yields. Furthermore, we found that using approximately six GGCMs and 10 GCMs was sufficient to determine the modelling uncertainty across the nine crop models and 32 climate models we considered, while a cluster-based selection of 3–4 GGCMs can effectively represent the full ensemble. These results highlight the importance of considering different ensemble configurations to better project crop yields and effectively utilize multiple models for specific applications.

**(4) Incorporating machine learning and crop model to predict crop yields and constraining the overall uncertainty.** This work presents a hybrid model that combines machine learning and crop models to predict maize and soybean yields while considering the impact of extreme climate events and crop pests and diseases (CPD). We found that our approach substantially increases the model performance compared to using only a crop model. The  $R^2$  values for maize and soybean were 0.05-0.32 and 0.05-0.32 for the crop

model, and 0.37-0.58 and 0.2-0.35 for the hybrid model, respectively. The nRMSE was around 0.2 for both crops in the hybrid model. Our analysis revealed that CPD, heat, and drought are the main predictors for soybean, while cold days, CPD, and drought are the main predictors for maize. Our hybrid model reduced the uncertainty around 48-67%. During T1 (2040-2069), the main source of uncertainty for maize was the crop model, while for soybean it was the general circulation model (GCM). During T2 (2070-2099), the shared socioeconomic pathway (SSP) became the main source of uncertainty. Our findings suggest that a hybrid approach that combines machine learning and crop models is an effective way to predict crop yields while considering the impact of various factors such as extreme climate events and CPD. This method can provide valuable insights for farmers, policymakers, and other stakeholders in the agricultural industry.

#### **(5) Climate drivers impact on global food security under future climate change.**

In this section, we first identify and compare the primary climate drivers affecting crop yields and their changes under climate change. Secondly, we robustly estimate the global and regional sensitivity of crop yields to climate drivers and quantify yield changes during strong oscillation phases in historical and future periods. Thirdly, we assess the significant areas impacted by strong oscillation phases. Finally, we analyze the contribution of different sources of uncertainty in the impact of climate drivers on global crop yields. For maize and wheat, ENSO is the dominant driver in Europe and Northern Asia, while IOD is the primary driver in Africa. However, NAO is expected to be the main driver influencing crop yields in most areas of the northern hemisphere from 2000 to 2099. Compared to historical data, NAO has a large impact on soybean and rice yields. During strong negative NAO and positive ENSO phases, maize, soybean, and rice yields decrease significantly. In contrast, soybean and rice yields increase during strong negative ENSO phases. During strong positive IOD phases, wheat and soybean yields decrease significantly in most areas. Extreme NAO significantly affects rice and soybean yields. We also analyze the uncertainty in this study and find that GCM-induced variance is higher than SSP and GGCM. However, the uncertainty contribution of NAO to SSP was greater than that of GCM in rice and maize. Our study improves understanding of climate variability and its impact on global breadbaskets under climate change, with potential implications for improving the resilience of the global food system.

**KEY WORDS:** machine learning; crop yield projections; climate change; global climate model; crop model; uncertainty analysis

---

## 目录

摘要 .....	I
ABSTRACT .....	V
<b>第一章 绪论</b> .....	<b>1</b>
1.1 研究背景、目的和意义 .....	1
1.2 国内外研究进展 .....	3
1.2.1 气候变化与粮食安全 .....	3
1.2.2 基于机器学习方法预测作物产量 .....	5
1.2.3 作物模型在气候变化对产量影响分析中的应用 .....	7
1.2.4 多模型集合预测以及不确定性分析 .....	8
1.2.5 气候驱动因子对产量的影响 .....	10
1.3 研究中存在的问题 .....	13
1.4 研究内容和技术路线 .....	13
1.4.1 研究内容 .....	13
1.4.2 技术路线 .....	15
<b>第二章 基于机器学习预测作物产量及其与环境变量的响应关系</b> .....	<b>17</b>
2.1 材料和方法 .....	17
2.1.1 研究区概况 .....	17
2.1.2 数据来源 .....	18
2.1.3 模型构建 .....	18
2.2 结果分析 .....	25
2.2.1 模型性能评价 .....	25
2.2.2 变量的相对重要性 .....	26
2.2.3 小麦产量对预报因子的响应 .....	27
2.2.4 不同子区域小麦的主要影响因子 .....	31
2.3 讨论 .....	32
2.3.1 模拟精度评价 .....	32
2.3.2 植被指数对产量预测的贡献 .....	33
2.3.3 气候因素与产量的关系 .....	36
2.3.4 研究的局限性 .....	37
2.4 小结 .....	37

第三章 建立动态产量早期预报模型并识别关键生长阶段	38
3.1 材料和方法	38
3.1.1 研究区介绍	38
3.1.2 数据介绍	39
3.1.3 模型构建	43
3.2 结果分析	45
3.2.1 模型预报精度	45
3.2.2 产量预报的最优预报时段	49
3.2.2 变量重要性	50
3.3 讨论	52
3.3.1 模型精度	52
3.3.2 小麦的产量预测	52
3.3.2 玉米的产量预测	53
3.3.4 水稻的产量预测	53
3.4 小结	54
第四章 基于多模型集合预测气候变化对作物产量的影响及不确定性分析	55
4.1 材料和方法	55
4.1.1 全球作物模型	55
4.1.2 气候数据	57
4.1.3 聚类分析	57
4.1.4 不确定性分析	59
4.2 结果分析	61
4.2.1 不同集合的产量预测	61
4.2.2 不确定性分析	69
4.2.3 模型数量对不确定性的影响	73
4.2.4 单一模型对整体不确定性的影响	83
4.3 讨论	88
4.3.1 预测的不确定性以及模型选择	88
4.3.2 不同的不确定性来源	88
4.3.3 本研究的局限性以及展望	89
4.4 小结	89
第五章 基于机器学习提高网格作物模型对产量的预测并约束不确定性	91
5.1 材料和方法	91

5.1.1 数据介绍 .....	91
5.1.2 NWAI-WG 统计降尺度方法 .....	93
5.1.3 GGCM 模拟器 .....	93
5.1.4 不确定性分析 .....	94
5.1.5 混合模型建模方式 .....	94
5.2 结果分析 .....	96
5.2.1 历史期间模拟精度 .....	96
5.2.2 产量与预报因子的响应关系 .....	97
5.2.3 气候变化对产量的影响 .....	99
5.3 讨论 .....	103
5.3.1 预报因子与产量的响应关系 .....	103
5.3.2 气候变化导致玉米和大豆减产 .....	104
5.3.3 机器学习约束不确定性 .....	105
5.3.4 不确定性来源 .....	105
5.3.5 研究的局限性 .....	106
5.4 小结 .....	106
<b>第六章 未来情景下气候驱动因子对作物产量的影响 .....</b>	<b>108</b>
6.1 材料和方法 .....	108
6.1.1 大气环流指数 .....	108
6.1.2 产量变化的主要影响因子 .....	110
6.1.3 敏感性分析 .....	110
6.2 结果分析 .....	112
6.2.1 未来 IOD、ENSO 和 NAO 的预估 .....	112
6.2.2 历史和未来产量变化及其与环流指数的显著性关系 .....	113
6.2.3 主导因子的变化 .....	114
6.2.4 强烈震荡阶段产量的变化 .....	118
6.2.5 产量对大气环流指数的敏感性 .....	121
6.2.6 研究的不确定性分析 .....	124
6.3 讨论 .....	126
6.3.1 气候变化导致对产量影响的主导因子发生了转变 .....	126
6.3.2 大气环流强震荡对产量的影响 .....	128
6.3.3 本研究的不确定性来源 .....	129
6.4 小结 .....	129

第七章 结论和展望 .....	131
7.1 主要结论 .....	131
7.2 创新点 .....	133
7.3 研究不足和展望 .....	133
参考文献 .....	135
附录 .....	156
致谢 .....	170
作者简介 .....	172

## 第一章 绪论

### 1.1 研究背景、目的和意义

随着人民生活水平提升，粮食需求在不断增加，世界粮农组织（FAO）一份报告表明，到 2050 年，全球粮食需求将增长 60%以上（FAO 2018）。然而在全球气候变化背景下，极端气候事件频繁发生，为粮食生产带来巨大威胁。因此，量化气候因素对作物生产力的影响成为了国内外普遍关注的重大科学问题。最近的 IPCC AR6 一份关于全球升温 1.5°C 的特别报告中指出（Mbow et al. 2019），粮食安全将越来越受到预计的未来气候变化的影响。在不同的社会经济排放情景（SSPs）中，全球作物和经济模型预计，由于气候变化（RCP 6.0），2050 年谷物价格将上涨 1–29%，鉴于越来越多的极端事件和相互影响，粮食系统受到破坏的风险正在增加。尽管过去几十年由于使用新品种和优化的农业措施可以提高作物产量，但是粮食生产在很大程度上仍然受到气候因素的影响，例如高温和干旱胁迫等。这些因素持续威胁着粮食安全，并且未来可能会变的更加严重。因此准确及时的作物产量预测对制定决定性的农业政策和保障粮食安全至关重要（Li et al. 2021a; 周营 2018; 车少静等 2005; 孙晶晶等 2018），特别是考虑到全球人口增长和生活水平提高导致粮食需求飙升（Tilman et al. 2011）。

通常，基于过程的作物模型和统计模型是用于产量预测的常用方法。基于过程的作物模型（例如 APSIM、CERES 和 STICS）可以模拟作物生长和产量形成过程，并能够分析作物产量与环境条件之间的相互作用（Brown et al. 2018; Feng et al. 2019; Peng et al. 2020b）。然而大规模逐个站点运行作物模型非常耗时，并且通常需要大量的田间观察数据（例如品种特征、管理选择和土壤特性）才能有效地校准模型。此外，与极端气候事件相关的过程在大多数作物模型中被大大简化，导致产量模拟的准确性降低（Li et al. 2019d; Peng et al. 2020a; Schwalbert et al. 2020）。与基于过程的作物模型相比，统计模型（如传统统计模型和机器学习方法）在大规模作物产量估算中效率更高，因此应用更广泛（Cao et al. 2021a; Peng et al. 2020a）。机器学习是一种先进的统计技术，可以分析预测变量和响应变量之间的层次关系和非线性关系（Besalatpour et al. 2014; Naimi et al. 2021; Zeraatpisheh et al. 2019; Zeraatpisheh et al. 2021）。它使用计算方法直接从数据中“学习”信息，而不依赖于预先确定的方程式作为模型。它已被广泛用于通过整合多个来源的环境数据来估计或预测作物产量（Anna et al. 2018; Cao et al. 2021a; Cao et al. 2021b; Li et al. 2021a）。然而，这些研究在产量预测中，一般都是基于完全的生长阶段来进行预测或模拟，如何准确的在收获期前进行产量预报，往往可以给决策者更多的时间来采取措施。在构建基于机器学习的产量预测模型是，识别出

哪些生育时段的变量起关键作用，对于分析产量的影响因子具有重要意义。此外，提高或优化机器学习以提高产量预测精度，对区域粮食安全也具有很大的价值。然而，机器学习也有一定的局限性。机器学习算法是基于数据的统计分析，无法直接反映作物与环境之间的生理生态过程，因此在研究气候变化对于作物产量的影响时，机器学习算法的解释能力和适用性都相对较差。

由于作物模型过于简化极端气候不能充分的考虑因子之间的相互关系，而机器学习可以考虑非线性响应关系，可以更好地预测不同气候条件下的作物生长情况。因此，基于过程模型和统计模型构建的混合模型已经广泛的被应用 (Feng et al. 2019; Li et al. 2021b)。建立混合模型可以考虑作物模型对生长过程的解释，也能考虑一些被过度简化的极端事件对产量的影响。有助于更好的预测气候变化对产量的影响，并探索极端事件对产量影响的新见解。然而，这样的方法在大尺度上应用需要很高的计算成本，因此只能在区域上或者选择相应的站点来模拟 (Feng et al. 2019; Li et al. 2021b)，很难直接在大尺度上应用。全球网格作物模型 (GGCM) 提供了空前的作物模型模拟，考虑了不同的模型结构以及假设 (Müller et al. 2019; Wang et al. 2021c)。Franke 等 (2020a) 基于全球网格作物模型比较 (GGCMI) 第 2 阶段工作的新参数扫描数据集开发了一组全球网格作物模型模拟器。该方法在计算上易于处理，同时结合了统计模型和作物模型的有点可以较好的捕捉气候变化对产量的响应，并得到了广泛的应用。例如 Zabel 等 (2021) 利用全球作物模型仿真器分析了玉米、水稻、大豆和小麦四种作物在未来气候情景下品种的适应对粮食生产的潜力。结果显示，在高变暖水平 (SSP585) 下，全球 39% 的农田可能需要新的作物品种，以避免到本世纪末气候变化造成的产量损失。而在低变暖水平 (SSP126) 下，85% 的当前耕地可以利用现有品种在农业生态区内转移以适应气候变化。然而，无论是在区域尺度还是全球尺度，作物模型预测产量有极大的不确定性 (Jägermeyr et al. 2021; Jiang et al. 2022; Müller et al. 2021; Tao et al. 2018; Wang et al. 2020b)。这些不确定性与模型结构，参数，气候模式 (GCM) 或作物模型等因素影响 (Wang et al. 2020b; Xiong et al. 2019)。

目前在不确定性分析方面，存在三个未被考虑的问题。第一，不同的模型组合对不确定性分析是否有影响。例如，如果选择了三个基于类似机理过程的作物模型和五个气候模型进行不确定性分析，那么结论可能会受到气候模型或作物模型的影响，具体情况因模型组合不同而异。因此，不同的模型组合在很大程度上影响了不确定性分析。本研究将试图探索一种方法来比较不同模型组合对不确定性的影响。第二，建立大规模的模型集合需要大量时间来计算。因此，需要确定使用多少模型才能全面反映模型的不确定性，并减少计算时间。第三，不同模型在不同区域或网格上对集合的不确定性贡献是不同的。如何量化不同模型在不同区域中的不确定性贡献也是需要解决的问题。解决这些科学问题有助于更好地理解不确定性，并为模型改进和选择提供参



考。

只有在充分考虑模型不确定性的情况下，才能可靠地预测气候变化对产量的影响。因为不同的模型假设和参数化方法可能会导致对同一气候变量的响应存在差异，甚至可能会导致完全相反的结果 (Swann et al. 2016; Wang et al. 2017b)。如果直接将所有模型平均来量化气候变化对产量的影响，可能会导致集合中抵消不同模型所描述的机理过程，从而几乎不会对产量造成影响。例如，Wang 等 (2020b) 基于 8 个作物模型和 28 个气候模式预测未来，新南威尔士州巴尔拉纳德和沃加沃加小麦产量的变化。虽然不同单一模型预测的产量变化相对较大，例如 APSIM 模型预测在 RCP8.5 情境下，在 2080s，巴尔拉纳德的产量将下降 31%，沃加沃加的产量将下降 59%。然而，在多模型平均中，产量变化仅为小于 5%。这表明多模型平均可能会抵消模型的特性，使得气候变化对产量的影响不明显。因此，更好地协调不同模型以提供可靠的预测对于制定适应策略、评估气候风险、维持全国和全球粮食安全具有重要意义。

我国人均耕地面积不足世界平均水平的 40% (孙晶晶等 2018)，且近年来，全球气候变化背景下，我国极端气候事件 (暴雨、干旱和高温等) 趋多趋强，气候风险水平呈上升趋势，加上我国幅员辽阔，气候变化对我国粮食生产的影响具有明显的区域差异性 (Li et al. 2019b; Piao et al. 2010; Yao et al. 2018)。因此，亟需针对不同气候分区，量化历史和未来的气候因素对主要粮食作物生长和生产力的影响，以增强我国农业生产对气候变化的适应性 (翟石艳等 2017; 吴普特 和 赵西宁 2010)。此外，我国的粮食需求增长也对全球粮食市场产生了巨大的影响。因此，将中国粮食生产与全球粮食安全联系起来，探讨气候变化对作物生产力的影响，具有重要的研究意义和现实意义 (宋承国 2009; 张宁宁等 2022; 朱晶等 2022)。同时，从全球尺度上讲，粮食生产和粮食安全也是至关重要的全球公共问题，特别是在人口和粮食需求迅速增加的情况下 (Lam et al. 2013; Wheeler and von Braun 2013)。最近，粮食系统变得更加全球化和相互依存，许多发展中国家从国际粮食贸易中受益，以减少饥荒 (Puma et al. 2015)。然而，气候变化和变化 (通常导致极端干旱、潮湿和热浪压力) 可能会影响粮食供应，并威胁整个粮食系统的稳定性 (Gohar and Cashman 2016; Ray et al. 2015)。这些风险可能同时暴露在多个粮食产区中，尤其是在最近几十年 (Gaupp et al. 2019)。因此，建立气候变化与产量的关系，可以提高对气候变化影响全球粮食产区的理解，同时对提高全球粮食系统的弹性具有潜在的价值。

## 1.2 国内外研究进展

### 1.2.1 气候变化与粮食安全

气候变化导致极端气候频繁发生：如洪涝灾害，极端热浪，干旱，霜冻等 (王劲松等 2012; 刘兴华等 2006)。一些研究发现极端高温呈现更加显著的增长趋势，然

而周期的显著性却降低,表明未来极端气温将呈现更加接近线性的增长(Li et al. 2019a; Wang et al. 2016a)。相反,霜冻等极端低温事件的趋势则显著下降。Yao 等(2019)基于全球气候模式(GCM)与统计降尺度模型预测了中国未来干旱变,发现中国整体未来干旱风险将会继续增大。一些子区域的干旱可能会有减缓,例如中国的东南部,东北等地区。虽然,未来降水也会有增加,但是由于未来极端降水出现的更加频繁,增加部分的降水可能很大一部分是极端降水,这样会导致旱涝灾害频繁发生,且短期的暴雨可能会加大水土流失,破坏农田等风险(Li et al. 2019b; Li et al. 2020a)。未来我国大部分地区极端降水会有 10-20 年的周期变化,CMIP6 和 CMIP5 模型都可以捕捉到历史时期极端降水的空间分布。然而,CMIP6 模型通常表现出更好的模拟指数年际变化的能力(Wang et al. 2022a)。对于温度和 CO<sub>2</sub> 来说,与 CMIP5 相比,CMIP6 预测的温度和二氧化碳浓度更高(Jägermeyr et al. 2021)。这将会对粮食生产产生很大的影响,并影响全球粮食安全。

气候变化对农业生成造成了严重的影响,通过多模型比较分析表明,全球变暖已经减缓了大多数小麦种植地的产量增长。整体上每增加 1°C,产量将下降 6%,这样的下降将随着温度升高变得更严重(Asseng et al. 2014)。虽然极强的降水也会导致产量减产,但是一些研究表明,极端降水其实对产量的影响是有限的,因为真正影响到产量的大暴雨其实出现的频率是非常少的,然而小雨(或者毛毛雨)则对产量有较大的负面影响,因为小雨出现的频率非常高,但是其水分还无法为作物提供有效的水分。在下小雨的时候往往伴随着低辐射和高的病虫害风险,因此小雨可能对产量的负面影响大雨暴雨(Lesk et al. 2020)。此外一些复合事件也对粮食减产有较大的潜在风险,如干热复合事件,湿热复合事件,土壤-大气干旱复合事件等(Heino et al. 2022; Lesk et al. 2021; Lesk et al. 2022; Zhou et al. 2019a)。Jägermeyr 等(2021)利用最新的 12 个全球网格作物模型以及 5 个气候模式计算了 CMIP6 和 CMIP5 中气候变化对产量的影响。结果表明,未来气候变化对产量影响的紧急约束事件出现将会提前。说明应该立即采取一些适应措施来减缓气候变化对农业的负面影响,因为目前已经十分紧迫(Peng and Guan 2021)。

已有很多研究探讨了不同措施对适应气候变化的效果。例如通过设计最新的品种来适应气候变化(Zhao et al. 2022)。一些研究表明未来强大的适应潜力取决于新的适应品种,到本世纪末,全球 39%(SSP585)的农田可能需要新的作物品种,以避免气候变化造成的产量损失,对于 SSP126 情景,85%的区域可以从现有品种中提取,以适应农业种植区内的变化(Zabel et al. 2021)。考虑到气候变化对不同区域的影响,未来很大一部分种植地区将不再适宜种植作物,种植带将向两极地转移(Franke et al. 2022; Kummu et al. 2021)。还有许多研究分析了不同的适应措施下产量的变化,如:通过调整播期(McDonald et al. 2022; Minoli et al. 2022),进行土壤肥力方面的管理以及投资,

(McCullough et al. 2022), 优化种植面积以及肥料的使用 (Liu et al. 2021b) 等。不同的适应方式会有不同的适应效果, 制定未来适应措施应充分考虑可持续性, 经济效益, 时效等来优化适应方案。

### 1.2.2 基于机器学习方法预测作物产量

统计模型主要包括线性模型和非线性模型。一些研究使用传统的回归模型进行产量预测作物产量或分析作物产量与气候因素之间的关系。例如Lobell等(2007)使用多元回归模型预测美国加利福尼亚州1998-2003年的作物产量, 发现使用具有 2-3 个气候变量的简单方程可以解释观察到的产量变化的三分之二以上。然而, 与非线性回归模型相比, 传统的统计模型(例如, 线性回归模型)通常显示出较低的准确性。实际上, 作物产量与多环境因素之间的关系通常是非线性的(Jeong et al. 2016; Li et al. 2007; Li et al. 2019d)。机器学习可以更好的考虑非线性的响应关系。最近, 许多研究开发了基于统计的作物产量预测模型, 如人工神经网络(ANN)、最小绝对收缩和选择算子回归(LASSO)、支持向量机(SVM)和随机森林(RF)等(Anna et al. 2018; Cao et al. 2020; Liakos et al. 2018; Norouzi et al. 2010; Peng et al. 2020a; 钟仁海 2022)。例如Leng and Hall(2020)使用传统的线性回归模型和随机森林模型预测美国1980-2010年的玉米产量变化。结果表明, 随机森林模型( $r = 0.93$ ,  $RMSE = 246$  kg/ha)比线性回归模型( $r = 0.51$ ,  $RMSE = 506$  kg/ha)模拟精度更高。Li等(2007)使用遥感数据、多元回归和人工神经网络(ANN)技术估算了美国中西部和太平洋地区的大豆和玉米产量。他们发现ANN( $r = 0.73-0.97$ 和 $RMSE = 518-1281$  kg/ha)比多元回归( $r = 0.58-0.93$ 和 $RMSE = 868-1681$  kg/ha)。Cai等(2019)将遥感和气候数据与机器学习方法相结合, 对澳大利亚的小麦产量进行了估算, 他们的方法具有良好的性能( $R^2 = 0.75$ )。在不同的机器学习方法中, 基于决策树的随机森林模型已被广泛应用于不同的研究领域(Rehfeldt et al. 2012; Singh et al. 2017; Zhao et al. 2019), 并且在产量预测方面有较好的表现(Han et al. 2020; Maya Gopal and Bhargavi 2019)。例如, Han等(2020)对比不同机器学习方法(支持向量机、高斯过程回归和随机森林), 他们发现机器学习在这三个模型中表现最好。此外, 机器学习能够识别每个预测变量对产量的非线性关系和相对重要性。

多源环境数据已被广泛用于预测作物产量(严海军等 2022; 张菊等 2021)。整合更多信息和预测因子, 例如土壤特性、极端气候指数和植被指数, 对于提高模型性能非常重要。作物产量受气候和土壤条件的显著影响(Alexandrov and Hoogenboom 2000; Chakrabarti et al. 2014; Wang et al. 2016b)。例如, 短时间日最高气温超过  $33^{\circ}\text{C}$  的极端高温事件会极大地影响灌浆早期的小麦和玉米粒数(Barlow et al. 2015; Dawson and Wardlaw 1989)。日最低气温小于  $0^{\circ}\text{C}$  的极端低温事件与作物开花期不育和籽粒败

育密切相关 (Barlow et al. 2015)。干旱和洪水也会显著影响作物产量。例如, 极端干旱会影响根系生长和结构, 并可能导致产量大幅下降 (Schwalbert et al. 2020); 洪水可直接毁坏农田, 还可能导致内涝, 这对土壤健康有害, 会导致产量显著下降 (Li et al. 2019c)。土壤特性被认为是影响作物生长的一个重要因素, 因为土壤的水和养分储存能力使它们能够在某些年份的不利条件下维持作物生长 (Folberth et al. 2016; Wang et al. 2018b)。因此, 这些土壤和气候变量可以提供有关作物潜在产量的必要信息, 并可作为预测作物产量的输入。Cao等 (2021a) 在产量预测中使用了土壤属性数据 (例如, 土壤质地、有机碳含量、pH 值等), 并对比了考虑土壤数据与不考虑土壤数据的区别。

此外, 基于遥感的植被指数数据由于容易获取且能较准确的反应作物的长势, 已经被广泛的应用与产量预测中 (靳彦华等 2013; 隋学艳等 2016; 张菊等 2021)。其中归一化差值植被指数 (Normalized Difference Vegetation Index, NDVI) 和增强植被指数 (Enhanced Vegetation Index, EVI) 是常用的两个植被指数。Chen等 (2018) 利用遥感数据推导作物物候和叶面积指数, 并将这两个参数同化到作物模型中, 以提高华北平原区域尺度上冬小麦产量估算的准确性。他们结果显示,  $R^2$  从 0.26 增加到 0.42, RMSE 从 1012 kg/ha 减少到 737 kg/ha。Wang等 (2020d) 开发了一个基于卫星的生物物理模型 (BEPS) 来推导中国东北地区水稻的实际产量及其空间格局。他们发现 BEPS 模型提供了该地区水稻产量的可靠估计, nRMSE 小于 20.0%。

然而这些传统植被指数可能无法解决混合像素问题或检测植物的光合功能变化 (竞霞等. 2021; 薛金儒 和 吕肖良 2022)。此外, 这两个指数对于温度并不敏感, 常常无法准确的反应高温对作物生理功能的影响 (Wang et al. 2020f)。与这两个传统的植被指数相比, 日光诱导叶绿素荧光 (Solar-Induced chlorophyll Fluorescence, SIF) 可以较好的反应植被生理现象, 以及他们对温度的响应 (Dechant et al. 2020a; Dechant et al. 2020b; Song et al. 2018)。然而目前大多数 SIF 产品的分辨率都较低, 有一些研究利用 GOME-2 和 OCO-2 的 SIF 产品来分析作物的生产力, 然而, 与目前 GOME2 的精度较差 (40×40km) 很难准确的反映出农田的状况。与 GOME2 相比, OCO2 具有较高的分辨率 (1.3×2.3km) 然而由于 OCO2 为条带状分布 (Bacour et al. 2019), 因此很难准确的匹配到田间的位置。Duveiller等 (2020) 基于 GOME2 数据通过统计降尺度方法开发了一套空间精度为 0.05°×0.05° 的 SIF 产品。然而降尺度的过程往往会造成误差的积累, 无法准确的反应 SIF 真实的大小, 因此在分析中可能无法正确的反应作物生理状况。目前时空精度比较高的 TROPOMI 产品被广泛应用, 但该产品的时间序列较短 (2017年11月-至今), 基于两年的数据估算气候变化对农业的影响缺乏代表性。与 SIF 相比, NIR<sub>v</sub> 即植被的近红外反射率 (NIR<sub>v</sub>), 由 MODIS 数据集计算为总场景 NIR 反射率 (NIRT) 和 NDVI 的乘积, 它具有较高的时空分辨率, 且可以在一定程度上反

应植被的光合作用 (Badgley et al. 2017)。此外, NIRv与SIF具有很强的相关性, 并能弄清背景亮度, 叶面积以及冠层深度中光合能力随深度的分布的混杂效应。它已被用来预测FLUXNET站点的生态系统光合作用, 并在月度到年度时间尺度上改善GPP的全球网格化估计 (Badgley et al. 2019)。Huang等 (2019) 基于NIRv找出最适生产温度结果与从涡流协方差通量塔测量得出的估计值相当, 这为将NIRv代理用于区域映射提供了支持。跨越不同的植被类型。而平均值据估计全球农田的温度为25°C。Wang等 (2020f) 基于NIRv提出了最适生产力的温度, 然后基于响应曲线改善提高作物模型。但是目前考虑到不同气候区域和不同作物这样的影响仍然不清楚, 此外, 植被光合指数是否受其他极端天气影响, 例如, 除了高温和辐射, 是否还受降雨, 干旱或者低温的影响。

统计模型的一个主要局限性是它们通常只提供简单的影响评估, 而不是提供在实施不同适应措施下的不同的生理约束, 如干旱胁迫、高温胁迫等。因此, 统计模型的结果有时在帮助有针对性地发展适应性实践方面可能是模糊的 (Feng et al. 2019)。此外, 由于机器学习无法考虑作物的生长发育过程等, 在分析气候变化对产量影响时具有较大的局限性。例如, 当机器学习模型遇到前所未有的极端气候事件时, 其预测能力可能会受到限制。

### 1.2.3 作物模型在气候变化对产量影响分析中的应用

基于过程的作物模型已经被广泛应用于产量估算和预测, 以解释当地的气象要素、环境因子、作物品种和不同管理措施等它们之间复杂的交互效应 (Chen et al. 2012; Chen et al. 2020b; Tao et al. 2017)。它们可以提供对生理过程的深刻理解, 并反映作物与环境变量之间相互作用的影响。例如 Chen 等 (2020b) 使用 DSSAT-CERES-Maize 模型估计了玉米的产量, 并取得了很好的效果 ( $nRMSE < 8.0\%$ )。在该研究中, 该模型成功地模拟了玉米的生长和发育, 并通过分析作物生长的关键环节来预测了产量 (提前 35 天)。在这种情况下, 作物模型提供了一种高度可靠的工具, 可用于作物管理决策的制定和实施。还有一些分析基于作物模型分析不同胁迫对作物产量的影响。例如 Liu 等 (2017a) 分析了生育后期高温胁迫对冬小麦的影响。Chen 等 (2020d) 利用 DSSAT 模型分析冬小麦受干旱的影响, 并对比了不同区域差异。

基于过程的作物模型不仅可以用于作物产量的预测, 还可以评估不同的适应策略, 为制定最优的气候变化适应措施提供指导 (Feng et al. 2022a; Kersebaum 2022)。例如, Minoli 等 (2022) 利用 LPJml 模型研究了调整作物的播期对全球作物产量的影响。他们发现, 在调整生长期的前提下, 作物的实际产量可以提高约 12%, 这有助于减少气候变化的负面影响, 并增强二氧化碳施肥的积极效应。此外, 此外, 作物模型还可以考虑一些缓解全球变暖增加可持续性的措施 (He et al. 2022; Lu et al. 2022)。例如,

He 等 (2022) 评估不同保护性农业措施, 发现残茬回田可以同时实现固碳稳产, 可作为缓解气候变化的有效措施之一。同样, Chen 等 (2019) 基于杨陵地区三年的实验以及 DNDC 模型预测了气候变化下不同覆盖措施对产量以及温室气体排放的影响, 结果表明秸秆覆盖可以在保证产量的同时尽可能的减少温室气体的排放。整体上, 基于过程的作物模型具有广泛的应用前景, 为气候变化适应和粮食安全提供了有力支持 (刘苇航等 2022; 彭慧文等 2022)。

目前大部分作物模型都是基于站点尺度来分析的, 然而如果想在区域或者全球上模拟, 将需要大量的观测数据来校准模型并需要较大的计算量。全球网格作物模型 (GGCM) 提供了空前的模拟, 并考虑了灌溉和雨养, 已有研究对比了 GGCM 与田间数据对温度的响应, 发现 GGCM 可以较好的反映出温度的影响 (Zhao et al. 2016a)。目前 GGCM 数据已经被广泛应用于分析气候变化影响、粮食安全、气候变化适应等科学问题 (Franke et al. 2022; Li et al. 2019c; Lyon et al. 2022; Minoli et al. 2022; Wang et al. 2021c)。然而, 无论是基于站点或是网格的作物模型模拟, 它们都过于简化了一些胁迫或者极端气候响应的过程, 例如高温热浪、极端干旱等。因此可能导致产量预测的不准确。例如, 大多数作物模型模拟高温对叶片衰老和茎碳水化合物积累和分布的影响, 而不是直接模拟生殖器官的损伤和流程。这增加了作物模型产量模拟时的不确定性 (Schauberger et al. 2017b)。

鉴于作物模型和统计模型都存在局限性, 将这两种模型进行整合, 以实现优势互补并探索新的发现 (张朝等 2020)。例如, Feng 等 (2020) 通过 APSIM 模型模拟多个生长阶段的特定指标 (生物量) 然后结合极端气候指数 (干旱、高温、和低温) 输入机器学习模型, 提高了 33% 的模拟精度。说明作物模型在模拟时没有充分考虑极端气候, 导致对产量存在高估。Li 等 (2021b) 结合作物模型 (MCWLA) 与机器学习方法 (随机森林模型) 优化了干旱风险评估, 并将其应用于农业保险。然而, 混合模型在处理大规模数据时可能需要消耗大量计算资源, 这可能会限制该方法在大尺度上的应用。

#### 1.2.4 多模型集合预测以及不确定性分析

目前, 由全球气候模型 (GCM) 输出的气候数据用来驱动物物模型是用来分析气候变化对产量影响的常用方法, 并用来制定气候变化的适应策略 (Asseng et al. 2013; Chapagain et al. 2022; Wang et al. 2018a)。一些研究使用单一作物模型分析气候变化对产量的影响 (Chen et al. 2020c; Jha et al. 2019; Yang et al. 2015)。例如基于作物模型预测结果选择最优的适应措施或者品种 (Huang et al. 2020; Wang et al. 2018a)。然而, 由于对许多复杂生物物理过程的不完善理解, 基于过程的作物模型具有很大的不确定性 (Dokoohaki et al. 2021)。这些不确定性可能无法反映可靠信息, 并进一步限制最



终用户、农民和决策者的应用 (Chapagain et al. 2022)。因此, 基于一个模型的模拟具有很大的偏差 (Asseng et al. 2013)。为了建立稳健的预测, 在作物产量模拟中开发了多作物模型和 GCM 的集成 (Asseng et al. 2013; Jägermeyr et al. 2021; Rosenzweig et al. 2014a)。

然而基于多模型的预测有着很大的不确定性 (孙擎等 2019; 杨伟才 和毛晓敏 2018), 因此需要了解产量变化中不确定性的主要来源, 以便制定减少总不确定性的策略 (Wang et al. 2020b)。研究表明, 气候-作物模型模拟方法对作物产量预测的不确定性主要与不完善的模型、输入数据、假设和过程表示有关 (Tao et al. 2018; Tao et al. 2020; Wang et al. 2020b)。在这些来源中, 输入的气候模式和作物模型都是一个重要的不确定性来源, 并影响着产量的预测。

一些研究量化了产量预测的不确定性来源, 他们选择一些站点并对比多个 GCM 和作物模型, 他们的结果都表明作物模型是主要的不确定性来源 (Asseng et al. 2013; Tao et al. 2018; Tao et al. 2020)。但是, 作物模型会不会总是主要的不确定性来源? Wang 等 (2020b) 对比了澳大利亚两个站点和中国的两个站点, 发现不确定性来源其实是因地而异的。对于澳大利亚新南威尔士州来说, 属于典型的地中海气候, 因此气候变异主导产量变化, 所以作物模型是主要的不确定性来源。然而对于中国的华北平原, 这边属于典型的温带季风气候, 四季分明, 这就导致了作物模型是该地区的主要的不确定性来源。因此, 不确定性来源还有很大的区域差异, 可能受不同的气候类型影响。

那么对于全球不同气候类型来说, 产量预测的不确定性来源是如何分布的呢? Xiong 等 (2019) 利用 3 个作物模型和 5 个 GCM 分析了全球冬小麦不同区域不确定性来源, 结果表明对于所有变暖的气候模式, 中高纬度地区的总体不确定性远大于低纬度地区。作物模型产生的不确定性大于其他来源的总和。Müller 等 (2021) 使用 9 个作物模型和 45 个 CMIP5 和 34 个 CMIP6 的 GCM 预测了未来小麦、玉米、水稻和大豆的全球产量变化, 并分析了相关的不确定性。在全球范围内, 小麦和玉米的不确定性主要由作物模型决定, 但大豆和水稻对气候预测更为敏感。具有非常相似的全球平均变暖的气候模型可能导致非常不同的总体影响, 因此气候模型的不确定性仍然是农业影响不确定性的重要贡献者。他们用了非常庞大的集合, 并全面评估了不确定性, 然而, 不同子区域整体的表现如何仍然需要分析。对于决策者来说, 可能某个区域整体的结果更有利于制定措施。

模型的数量对于不确定性仍然有很大的影响, 如果可以选择出一定数量的模型并反映出整体的不确定性, 则可以极大的减少计算量 (Lee et al. 2021)。目前这样的分析在气候模式以及水文模型中应用比较多。有研究表明, 选择出 5 个比较典型的 GCM 可以基本上覆盖全部集合的范围 (Ruane and McDermid 2017)。对于水文模型, 有相

关研究分析了不同 GCM 数量下通过水文模型模拟的不确定性，他们发现，至少随机使用 10 个 GCM 可以有效的反应出整体的不确定性 (Wang et al. 2020c)。然而，在农业方面这样的分析较少，而在保证反映整体差异的同时减少模型数量可以降低模型的冗余。一些研究对比了不同模型数量集合对历史产量再现的能力 (Falconnier et al. 2020)。他们使用了 25 个作物模型，发现实测与模拟籽粒产量的差异随着集成模型数量的增加而迅速减小，至少随机选择 8 个模型可以得到比较满意的模拟结果。Asseng 等 (2013) 在站点尺度上发现随机选择 5 个作物模型基本上可以反映小麦产量预测整体的不确定性。然而这在不同区域，以及对于不同作物来说仍然有很大差异。此外，上述的这些研究是随机选择模型的，那么是否可以提出一些方案，来进行模型选择，以进一步降低模型数量。因此，进一步找出模型不确定性最小数目，可以对气候变化对产量影响方面的研究中模型的选择提供参考。

除了模型整体的数量，一些偏差较大的单一模型也会对模型的集合带来很大的不确定性。一些研究模拟了全球的产量，然后分析某一个模型对于全球整体产量预测不确定性的贡献 (Müller et al. 2021)。这虽然对于模型选择可以提供参考，但是每个模型带来的不确定性在不同区域上是不同的。因此，也有一些研究选择了一些站点，并分析单一模型对于这个站点不确定性的贡献。他们发现每个模型带来多大的不确定性也是因地而异的 (Wang et al. 2020b)。然而，这仍然无法对于全球不同地区提供参考。分析全球不同区域，单一模型带来的不确定性，对不同地区模型的应用、选择以及改进有很大的意义。

除了上述的一些问题还发现，鉴于单一模型有较大偏差目前大多数研究都倾向于用多模型平均。有一个比较认可的结论是，单一模型确实无法准确的反映出较为可靠的信息。但是多模型平均仍然存在一个问题：由于他们之间的响应差异，多模型平均可能会导致不同模型间的特性相互抵消，最终导致结果变化很小 (Swann et al. 2016)。因此，可能需要建立不同的子集合，即考虑了模型的集合，保证不会有太大偏差，也可以尽可能的规避模型之间相互抵消的情况。此外，对于模型不确定性来说也是一样的。如果选择的作物模型恰好比较相似，那么 GCM 可能就会成为不确定性的主导因子，反之亦然。所以模型的选择对于不确定性分析其实也是有很大的影响的。然而，目前很少有研究对比过这样的分析。本研究将回答这些科学问题，为模型不确定性，以及模型选择提供有用信息。

### 1.2.5 气候驱动因子对产量的影响

大气环流因子会影响一些局部地区的气候变化，与气候系统有着密切的联系 (段安民 和 张萍 2022; 郭品文 和 吴清传 2022)。一些气候驱动因子例如厄尔尼诺南方涛动 (ENSO) 会影响全球范围的天气变化 (陈浩 et al. 2023; 莫嘉琪 et al. 2006)。



例如, 在 ENSO 循环的异常导致太平洋中部和东亚之间存在遥相关。导致从中国南部向东北延伸至日本东部(黑潮延伸)的东亚极地锋面异常潮湿(干燥)(Wang et al. 2000)。在厄尔尼诺年份中, 在印度尼西亚地区的降雨量趋于减少, 而在热带太平洋中部和东部的降雨量增加。通常沿赤道从东向西吹的低层地面风(“东风”)反而减弱, 或者在某些情况下开始向另一个方向吹(从西向东或“西风”)。一般来说, 海洋温度异常越暖, 厄尔尼诺就越强(Cai et al. 2020)。拉尼娜事件则与厄尔尼诺相反。除了 ENSO 外, 北大西洋涛动(NAO)与印度洋偶极子(IOD)也对全球天气有着很明显的影响或调节作用(Jones et al. 1997; Saji et al. 1999; Saji and Yamagata 2003; Stephenson et al. 2006)。这些也会对全球生态、社会经济、农业生产等产生影响。

已有大量的文章分析了 ENSO 对不同方面产生的影响。例如 ENSO 会通过遥相关影响全球的热带农业地区增加了粮食不安全, 进而间接的影响粮食不安全人口中的儿童的影响健康(Anttila-Hughes et al. 2021)。此外 ENSO 还与一些地区冲突、羊群的死亡率、碳汇等有很大的相关性(Hsiang et al. 2011; Islam et al. 2004; Le et al. 2021)。例如, Le 等(2021)发现 ENSO 对亚洲、大洋洲和美洲许多地区的碳通量和陆地碳储存具有显著的影响, 在未来, 受 ENSO 影响的地区有所增加。他们还强调与季节性碳通量的响应相比, 季节性碳储量对 ENSO 的响应更强。虽然 ENSO 是认为的目前对全球气候影响最大的指数, 但是如果在分析的时候只考虑 ENSO, 仍然会漏掉一些很重要的信息。一个比较典型的例子就是 2019 年澳洲火灾, ENSO 在 0.5 左右, 而 IOD 是在一个极端正向的相位(Liu et al. 2022; Phillips and Nogrady 2020)。正相位的 IOD 起的赤道信风东移的干燥效应可以促进火灾条件并导致特大野火(Saji and Yamagata 2003)。此外, NAO 也对北半球的气候有显著的影响。例如 NAO 在欧洲驱动着温度、降水、积雪和风的模式和极端情况并对该地区的碳汇有着十分明显的调节作用(Bastos et al. 2016)。

在农业方面, 气候驱动因子对作物产量也有很大影响。相比于基于气象要素预测产量, 将 ENSO 考虑进去, 可以大幅度的提高提前预测的时间(Iizumi et al. 2013; Iizumi et al. 2014)。后续的一些研究考虑了更多的指数(IOD, NAO, ENSO)结合作物模型, 发现全球有大约三分之二的地区受气候变异影响(Heino et al. 2018)。考虑到不同模型有差异, 有研究用了不同的作物模型, 且考虑了不同的适应措施结果表明在雨养和充分施肥的情况下, 可以观察到作物产量对气候变异有强的敏感性, 而在充分灌溉的情况下, 这种敏感性往往较低(Heino et al. 2020)。Anderson 等(2019)量化了气候变异模式如何影响作物产量, 他们发现厄尔尼诺南方涛动(ENSO)、印度洋偶极子(IOD)、热带大西洋变率(TAV)和北大西洋涛动(NAO)共同占全球玉米、大豆和小麦总量的 18%、7% 和 6% 的变异性。由于 ENSO、IOD 和 NAO 可以提前预测作物

产量，并有较好的效果。因此更好地理解作物生产与这些气候变异之间的关系可以提高全球粮食系统对气候相关冲击的抵御能力。

也有一些研究在区域上分析气候变异对产量的影响。例如美国作物减产风险的变化与大西洋和太平洋的大规模气候振荡有关 (Schillerberg and Tian 2020)。然而这些研究都是用线性模型分析的，而气候震荡与作物产量的关系可能是非线性的。Wang 等 (2020a) 基于机器学习量化了 ENSO 对澳洲冬小麦产量的影响。结果表明，可以在收获开始前三个月的提前期预测州一级的小麦产量变化。此外，南方涛动指数 (SOI) 是估算冬季降雨主导州作物产量的最重要因素。

气候变化往往会改变海洋表面温度 (SST) 或海平面压力 (SLP) 的变化格局 (Collins et al. 2010; Hori and Ueda 2006; Vecchi and Soden 2007)，从而扩大了 ENSO 等气候驱动因素的可变性。例如，根据 IPCC 第六次评估报告 (AR6) 的四种可能的排放情景，预计 ENSO 区域的 SST 变异性将大幅增加 (Cai et al. 2022)。最近的研究表明，在高排放代表浓度路径下 (SSP585)，预计 ENSO 与干旱之间的关系更强 (Singh et al. 2022)。这些变化也可能影响全球粮食生产。了解历史和未来受强烈振荡阶段显著影响的地区，有助于农民和决策者有更多时间来应对不同的不利条件。然而，目前还没有任何研究调查未来气候驱动因素与未来全球作物产量之间的关系。此外，以前的大多数研究都是通过线性模型将作物产量和气候波动联系起来的，这并不能说明它们之间潜在的非线性关系。

如今粮食系统变得更加全球化和相互依存，许多发展中国家从国际粮食贸易中受益，以减少饥荒。由于有多个粮食产区因气候变异而减产的风险非常高，识别主要气候驱动因素可以提高对气候变化与粮食安全之间关系的理解，尤其是在这些气候风险较大的地区 (Feng et al. 2022b; King et al. 2020)。尽管目前已经有一些研究量化了气候变异对产量的影响，但是这个影响是不是一成不变的呢？Feng 等 (2022b) 发现气候驱动因素的影响不是固定不变的，自上世纪以来，澳大利亚小麦产量的主要气候驱动因素发生了巨大变化，这也预示着全球范围内的潜在变化。然而，研究对多气候驱动因素对全球作物产量及其非线性响应的变化影响仍然认识不足，特别是在未来气候变化的情况下。

大气环流指数的预测和作物产量的预测，仍然存在很大的不确定性 (Jägermeyr et al. 2021; Lee et al. 2021; McKenna and Maycock 2021; Müller et al. 2021)。因此，它可能会对气候波动影响全球作物产量的分析产生较大的影响，这很大程度上增加了本研究的不确定性。了解此类分析中不确定性的主要来源可以从制定策略中受益，以减少总体不确定性，并提高粮食风险预测的稳健性。在作物产量变化预测中，大多数先前的研究发现，作物模型是不确定性的来源。然而，没有研究量化了气候模式-作物产量关联分析的不确定性。特别是由于气候波动预测和全球作物产量预测，这种不确定

性可能会被放大。本研究将解决上述存在的问题，以提高对气候变化对全球粮食产区冲击的理解，这对提高全球粮食系统的抵御能力具有潜在意义

### 1.3 研究中存在的问题

基于过去对产量预测以及气候变化影响分析方面已取得的结果，目前还存在以下几个问题有待进一步挖掘：

(1) 目前已有一些研究构建了基于机器学习的产量预测模型 (Jiang et al. 2020; Wang et al. 2020g)，然而大多数研究都没有探究环境变量对于产量预测的潜在影响，例如一些新的植被指数，作物光合替代指数。虽然一些研究基于多源线性回归分析了气候因子与产量的关系 (Li et al. 2020b; Lin et al. 2021)，但很多研究没有考虑极端气候的影响。过去基于线性模型的分析也很难解释气候因子和极端气候指数与产量之间的非线性关系。

(2) 过去的研究大多只能在生育期结束后模拟产量，无法提前预报作物产量，这对于农民和相关机构来说是一个挑战。他们需要提前采取措施以应对潜在的产量下降或灾害风险，保护作物和农民的利益变得更加困难。尽管有一些研究可以提前一个月左右预报产量 (Cai et al. 2019; Cao et al. 2021a)，但是这些研究没有考虑到特定生长发育阶段的影响。因此，研究需要建立基于特定生长阶段的产量预测模型，以提前预报作物产量并更好地为农业生产提供指导。

(3) 许多研究基于气候模式来驱动物物模型 (或多个作物模型) 以预测气候变化对产量的影响。然而，作物模型过于简化了作物生长对于极端气候的响应 (Rosenzweig et al. 2014a; Xiong et al. 2019)，另外，目前大多数作物模型都没有考虑病虫害对产量的影响。此外，作物模型和气候模式都存在较大的不确定性。虽然一些研究分析了产量预测的不确定性来源 (Jägermeyr et al. 2021; Tao et al. 2018; Wang et al. 2020b)，但是仍不清楚不同模型间的组合以及模型数量如何影响模型的不确定性。

(4) 直接分析气候驱动因子与产量的关系可以更早地预测产量的变化并解释产量波动的机制 (Anderson et al. 2023; Cao et al. 2023; Iizumi et al. 2014)。然而大多数研究是基于线性模型的，无法解释气候驱动因子与产量之间的非线性关系 (Anderson et al. 2019; Anderson et al. 2020; Heino et al. 2018)。气候变化以及温室效应加剧可能会影响气候驱动因子的变化规律 (例如 ENSO、IOD 和 NAO)。然而目前尚不清楚未来气候变化下这些驱动因子与作物产量之间的关系。

## 1.4 研究内容和技术路线

### 1.4.1 研究内容

(1) 利用多源环境变量建立基于机器学习的产量估算模型并分析非线性响应关系

基于气象数据、生育期数据、土壤属性数据、不同的植被指数 (EVI、NDVI、NIRv)、极端气候指数和作物产量数据采用随机森林模型 (RF) 和支持向量机 (SVM) 预测冬小麦产量并对比不同模型和不同区域的模拟精度。然后找出他们的主要预报因子, 并量化他们的重要性。最后分析预报因子与产量的非线性关系。

(2) 构建产量预报模型并识别关键生长阶段

基于多源环境变量以及农业调研数据 (冬小麦、春玉米、夏玉米、早稻、一季稻、晚稻) 结合随机森林模型构建基于机器学习的产量预报模型。对比不同预报时段模拟的精度, 找出关键生育阶段。对比中国不同区域模拟精度以及最有预报时段的差异。分析气候变化对不同生育阶段的影响。

(3) 基于作物模型分析未来气候变化对产量的影响及不确定性分析

基于 9 个作物模型和 32 个 GCM 预测全球小麦、玉米、水稻、大豆的产量变化。对比考虑二氧化碳施肥效应以及不考虑二氧化碳施肥效应的产量差别。基于聚类分析对全国粮食产地进行区域划分, 划分为 12 个子区域。基于另一种聚类分析对不同的预测结果进行聚类。分析不同模型组合的不确定性。基于不确定性分析方法 (ANOVA) 分析全球不同尺度 (网格、区域、纬度) 的不确定性来源。然后分析不同模型数量对不确定性影响, 找出最少有效数量。探讨不同区域单一模型对产量预测不确定性的贡献。

(4) 建立基于机器学习与作物模型的混合模型提高极端气候下的作物产量预测

基于 9 个作物模型、12 个 GCM、随机森林模型、极端气候指数、病虫害数据建立作物模型结合机器学习的混合模型来估算中国未来作物产量变化 (本研究以玉米和大豆为例)。模拟 SSP126 和 SSP585 情境下玉米和大豆的产量变化, 对比作物模型和混合模型预测的差别。分析预测因子与产量的非线性响应关系。然后对该研究的不确定性来源进行量化。

(5) 利用统计模型集合作物模型的方法分析大气环流因子对作物产量的影响

基于 5 个 GCM 模型计算历史和未来的 ENSO、NAO、IOD (1900-2099), 并对比历史和未来的变异性。通过机器学习结合作物模型找出历史和未来大气环流因子对不同作物 (麦、玉米、水稻、大豆) 产量影响的主导因子, 并量化这些指数处于异常相位对产量的影响。找出那些情况可能会导致全球多个粮食产区同时收到冲击。通过动态线性模型分析作物产量对 ENSO、IOD、NAO 历史和未来的敏感性变化。然后分析该研究所存在的不确定性。

需要注意的是, 本论文试图尽可能的考虑不同的研究区。但是结合机器学习预测产量, 需要大量的观测数据来训练模型。因此, 受可收集的数据影响, 各章节研究范围有所不同。其中, 第二章和第三章基于机器学习分析历史期间气候变化对我国作物

产量的影响。第四章为了充分考虑气候变化对不同气候类型和地区产量预测以及不确定性的影响，将研究区扩展至全球。第五章提出了利用机器学习约束作物模型不确定性的方法，但是仍然受所收集的数据限制，只分析了中国区域。第六章是机器学习结合作物模型方法来分析大气环流因子对产量的影响。鉴于大气环流因子的影响往往是全球性的，因此本研究在全球的尺度上探讨了未来气候变化下大气环流因子对粮食生产的影响。

#### 1.4.2 技术路线

本文基于气象数据、农业数据、遥感数据、土壤数据、气候模式数据结合作物模型和机器学习从不同尺度模拟和预测了作物产量。并分析了不确定性以及全球粮食生产的气候风险。主要包括：（1）基于机器学习以及多源环境变量实现产量模拟；（2）实现了收获期前的产量预报，可以提前 1-3 个月得到比较可靠的预报；（3）提出了一个新的框架基于多模型集合分析了未来气候变化对不同作物产量的影响以及预测的不确定性；（4）结合机器学习与环境变量（极端气候指数与病虫害风险）提高作物模型对产量的预测；（5）利用机器学习结合作物模型的方法分析大气环流因子对粮食安全的影响。具体技术路线如图 1-1 所示：

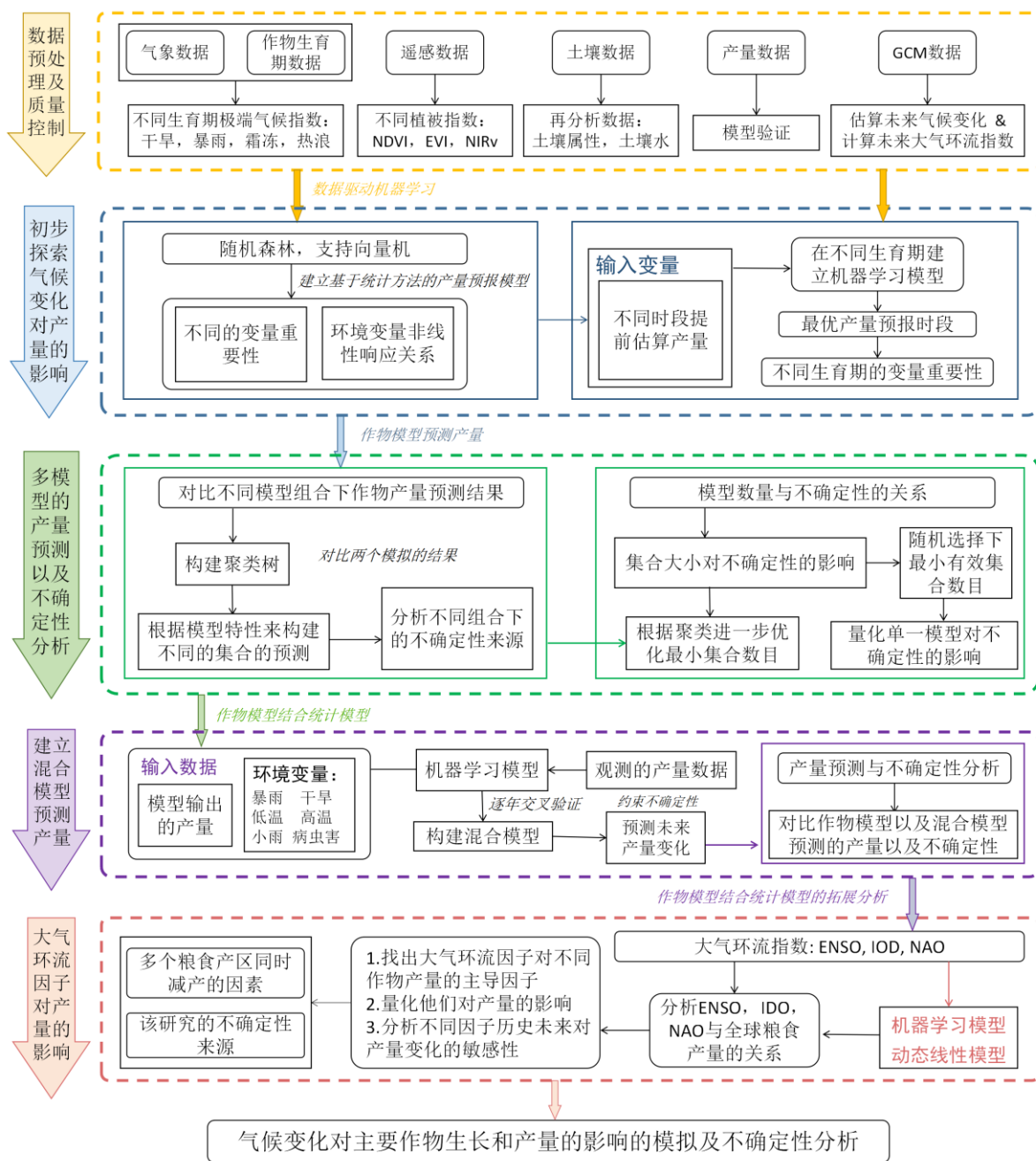


图 1-1 本研究技术路线图

Fig.1-1 Technical route of this study

## 第二章 基于机器学习预测作物产量及其与环境变量的响应关系

作物产量受多种环境因子影响，包括特定地点的土壤特性、气候变量、光合作用等因素。因此，准确、及时地预测作物产量对于保障国家粮食安全至关重要。然而，不同环境变量的选择和机器学习模型的应用，会对产量预测精度产生重要影响。例如，不同气象数据源、基于遥感的植被指数（如 NDVI、EVI、NIRv 等）和土壤特性数据的使用等。此外，不同环境变量（例如不同极端气候指数）在不同生育阶段对产量的非线性响应关系仍需深入研究。因此，本研究以小麦为例，探索了不同植被指数在小麦产量预测中的应用潜力，并分析了主要的预测因子与小麦产量之间的非线性关系。本章的主要目标是：（1）基于多种环境变量和机器学习模型，预测中国小麦产量；（2）量化预测因子的重要性；（3）分析预测因子对我国小麦产量变化的影响。

### 2.1 材料和方法

#### 2.1.1 研究区概况

本研究主要包括中国三个气候分区：西北荒漠地区（I）、华北湿润半湿润温带地区（II）、华中华南湿润亚热带地区（III）。中国小麦种植面积总计  $2.3 \times 10^7$  公顷。由于我国幅员辽阔、地形复杂（Piao et al. 2010），不同子区域的实际环境条件不同，对作物的影响也不同。因此，为了更好地预测小麦产量，根据地理和气候条件将小麦种植区划分为三个子区域（赵松乔 1983）。

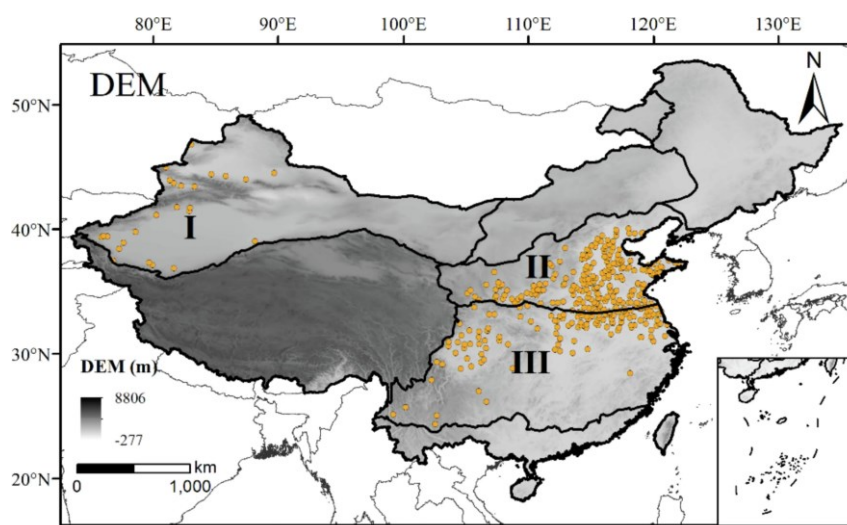


图 2-1 中国高程小麦站点三个子区域的空间分布

Fig.2-1 Map showing the spatial distribution of study sites (orange dots) and elevation (gray shading) for three sub-regions in the wheat-growing area of China.

### 2.1.2 数据来源

本研究收集了全国 373 个站点的小麦实验数据。其中包括 131 个农业气象站（2000-2013）和 242 个国家农情监测站（2014-2020）的数据。由于两种数据来源在管理实践和收获方式上基本一致，因此将两组数据合并，获得了 2001-2020 年的小麦产量数据。在数据获取中，本研究排除了产量超出当期报告范围以及落在平均值 $\pm 1.5$  IQR 之外的数据，共使用了 1936 条数据。该数据覆盖小麦生长季节的九个阶段：播种、出苗、分蘖、越冬、返青、拔节、抽穗、乳熟和成熟期。为了更好地解释预测因子在不同生长阶段对作物产量的影响，将小麦生长季节分为四个主要生长期大约是：T1：播种-分蘖期（11 月-次年 3 月）；T2：分蘖-拔节期（12 月-次年 3 月）；T3：拔节-抽穗期（3 月-4 月）；T4：抽穗-成熟期（5 月-6 月）。

基于遥感的植被指数可用于监测作物生长状况。本研究选择 2001-2020 年期间的 NDVI、EVI 和  $NIR_v$  作为作物产量的预测因子。 $NIR_v$  被认为是可以反应光合作用的代替产品，计算为 NIR 反射率和 NDVI 的乘积 (Badgley et al. 2017)。NDVI、EVI 和 NIR 反射率来自 16 天的全球植被指数产品，空间分辨率为 500 米 (MOD13A1 V6)。所有的植被指数都从 Google 地球引擎 (GEE) 平台提取至站点，然后整合为每个站点四个生长阶段中每个阶段的平均值。

土壤特性也在决定作物产量方面发挥作用。本研究的工作使用三个土壤特征作为预测因子：土壤有机碳含量 (SOC)、土壤容重 (SBD) 和粘土的阳离子交换能力 (CLAY)。这些土壤特性数据来自中国土壤粒度分布数据集 (<http://globalechange.bnu.edu.cn>)。

本研究使用的气候相关预测因子包括气候变量和极端气候指数 (ECE)。气候变量是四个生长阶段 (T1-T4) 的温度 (Tmean)、降水量 (Prp)、风速 (Ws)、相对湿度 (RHum) 和日照时数 (用于计算太阳辐射, Rad) 的平均值。这些数据来自中国气象数据共享网 (<http://data.cma.cn/>)。本研究计算了四个小麦生育期的四种 ECE，以反映高温、霜冻、干旱和极端降水。高温 (28°C) 和霜冻 (0°C) 事件分别是温度高于和低于固定温度阈值的天数。标准化降水和蒸散指数 (SPEI) 用于调查四个生长时期的干旱强度。SPEI 是通过标准化降水量和参考作物蒸散量 ( $ET_0$ ) 之间的差异来估算的。为了更好地地区分每个生长期的干旱强度，SPEI 以 1 月的时间尺度计算。作物产量也受到降水或极端降水的强烈影响。小雨和强降水对作物产量的影响不同。因此为了研究不同降水强度对小麦产量的影响，将小雨定为 R5 (降水量在 0.1-5mm 之间的天数)，将强降水定为 R20 (降水量大于或等于 20mm 的天数)。

### 2.1.3 模型构建



本章的技术路线图如图 2-2 所示。首先使用物候数据将小麦生长季节分为四个时期。然后，使用两种机器学习方法（支持向量机和随机森林模型）预测小麦产量。还确定了每个预测变量的特征重要性以及预测变量和产量之间的非线性关系。

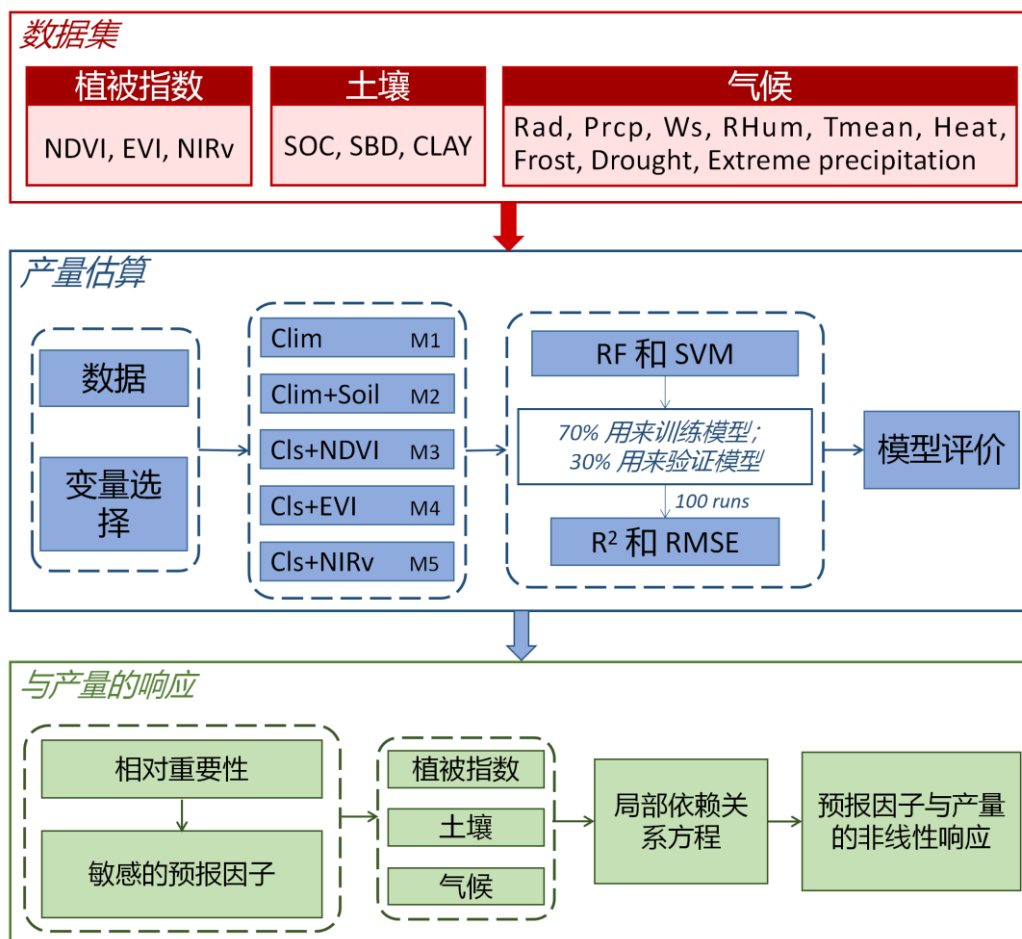


图 2-2 结合多源环境变量和不同机器学习方法的小麦产量预测模型框架

Fig.2-2 Framework of the wheat yield prediction model integrating multi-source data with different machine learning techniques.

注：图中 NDVI：归一化差异植被指数；EVI：增强型植被指数；NIRv：陆地植被的近红外反射率；SOC：土壤有机碳含量；SBD，土壤容重；CLAY，粘土的阳离子交换容量；Rad：太阳辐射；Prpc：降水；Ws：风速；RHum：相对湿度；Tmean：平均温度；DHW：干热风；Heat：高温事件；Frost：霜冻；Drought：干旱（SPEI）；Extreme precipitation：两种极端降水事件（R5 和 R30）；ML，机器学习（随机森林和支持向量机）；CLS：气候数据和土壤特性数据。

### 2.1.3.1 变量选择

本研究的输入数据由四个生长期的不同变量组成，这导致数据集较大。此外，很多变量可能会与产量有不合理的相关性，导致模型出现过拟合。本研究对变量进行了筛选：去掉了不显著的变量（ $P > 0.05$ ），如图 2-3 至图 2-6 所示。

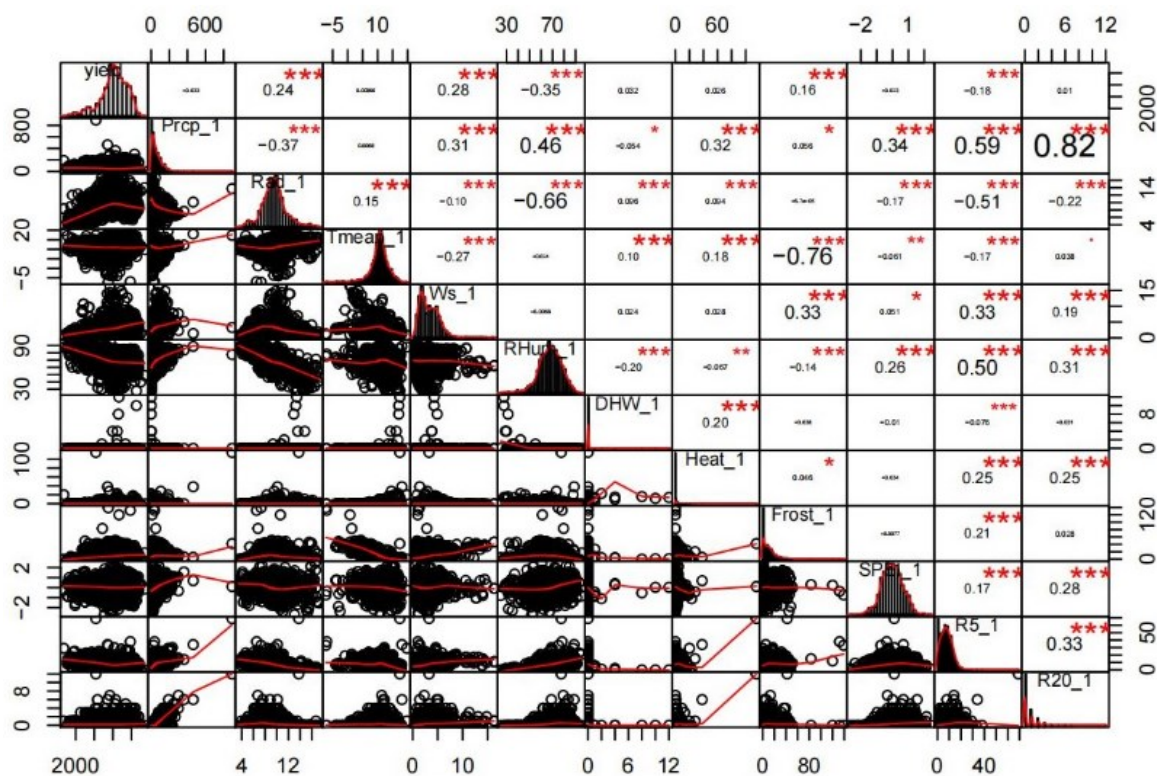


图 2-3 T1 时期小麦产量与气候相关预测因子的相关矩阵

Fig. 2-3 Correlation matrix of wheat yield and climate-related predictors during T1.

注: \*  $P < 0.05$ ; \*\*  $P < 0.01$ ; \*\*\*  $P < 0.001$

气候（气候变量和 ECEs）与产量的相关性和显著性水平随四个生育期的变化而变化。例如，在 T2 期间，Rad、Tmean、Ws、RHum、Frost 和 R5 与作物产量表现出很强的显著相关性（ $P < 0.001$ ），而 Prcp、DHW（干热风）和 Heat 则表现出不显著的相关性（图 2-4）。这主要是因为这一时期很少出现 DHW 和 Heat 等 ECE。

经分析发现，T4 的 DHW 和 Heat 与小麦产量呈显著相关关系（见图 2-6）。降水量在 T3 和 T4 期间与产量也有显著相关性，但在 T1 和 T2 期间则无显著相关性。值得注意的是，降水相关指数（R5）与整个生长季节（T1-T4）的作物产量也呈显著相关性，这表明在作物产量估算中定义 ECEs 非常必要。研究还发现，相比大雨（R20），小雨（R5）对产量的负面影响更为严重。此外，同样的 ECEs 在不同的生长阶段可能对小麦产量的影响有所不同。例如，在 T1-T2 期间，霜冻与小麦产量呈正相关关系，但在 T3-T4 期间则呈负相关关系。因此，在预测作物产量时，必须考虑关键的生长阶段或生长期。

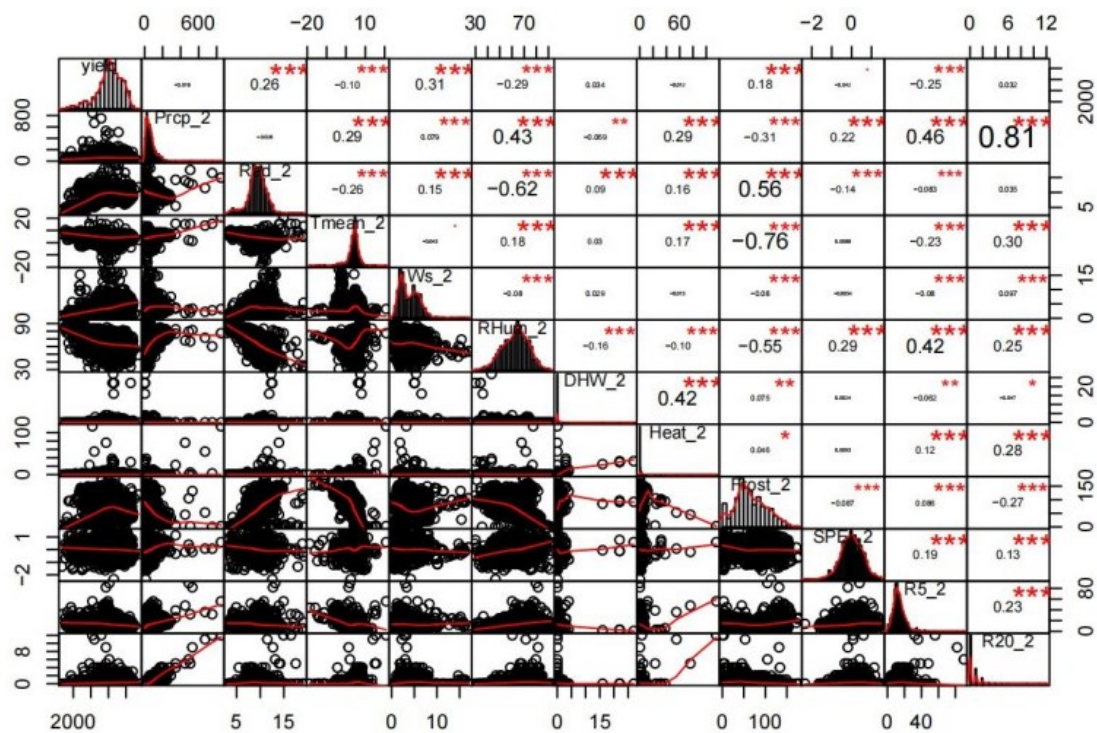


图 2-4 T2 时期小麦产量与气候相关预测因子的相关矩阵

Fig. 2-4 Correlation matrix of wheat yield and climate-related predictors during T2

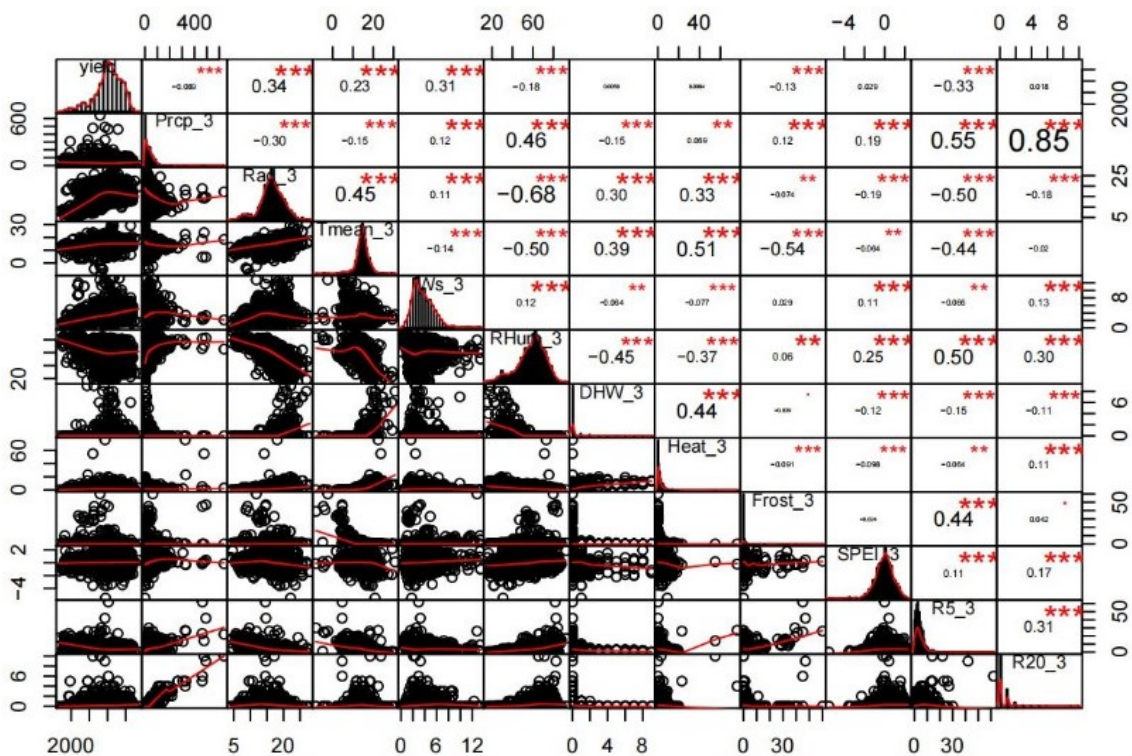


图 2-5 T3 时期小麦产量与气候相关预测因子的相关矩阵

Fig. 2-5 Correlation matrix of wheat yield and climate-related predictors during T3



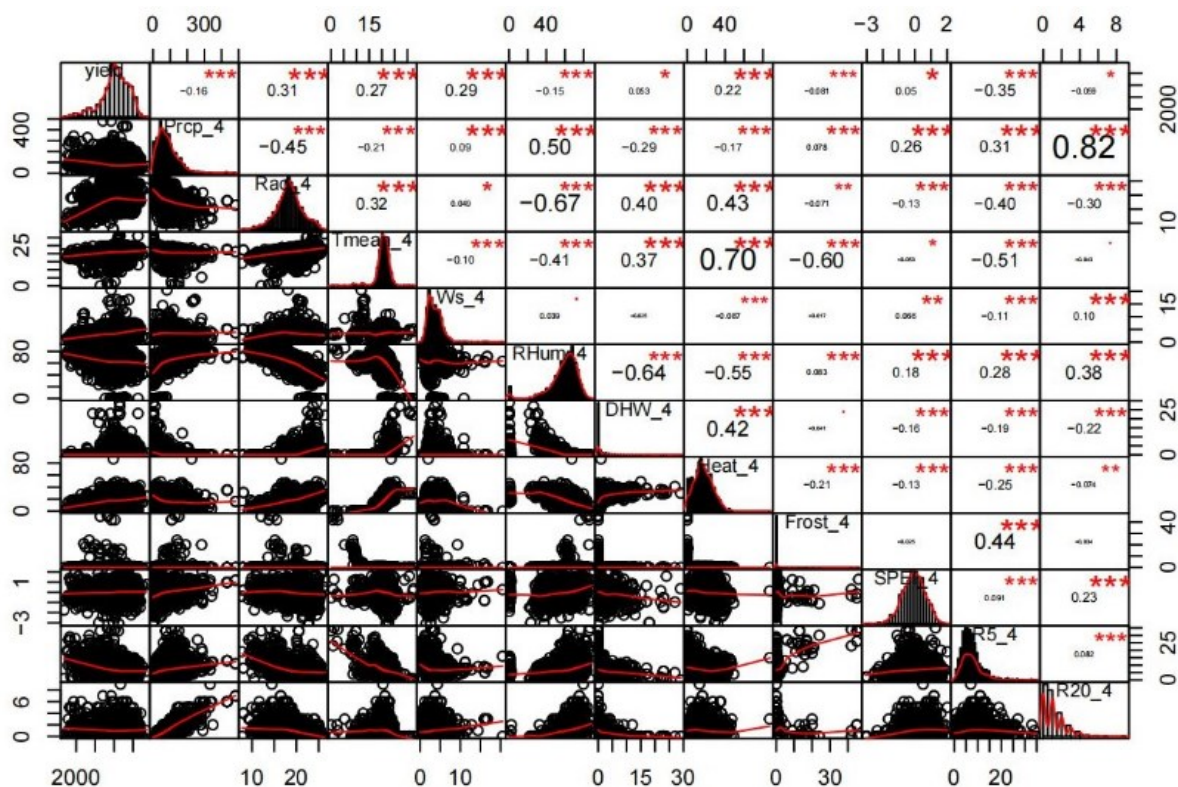


图 2-6 T4 时期小麦产量与气候相关预测因子的相关矩阵

Fig. 2-6 Correlation matrix of wheat yield and climate-related predictors during T4

### 2.1.3.2 支持向量机

支持向量机(SVM)是一种被广泛应用的机器学习模型,可用于回归和分类分析。相较于其他试图最小化观察数据和模拟数据之间的误差的机器学习模型, SVM 通过找到阈值内的最佳直线来拟合数据。在 SVM 模型中, 给定输入和输出数据的一组观察样本, 里面的最佳拟合线是具有最大点数的超平面 (Besalatpour et al. 2012)。

本研究使用了 SVM 的径向基函数 (RBF) 内核, 这种内核有两个参数: Cost 和 Sigma。在本研究中, Cost 的取值在 2、4、8、16、32、64 之间, Sigma 的取值在 0.02 和 0.08 之间, 间隔为 0.005。本研究采用了三重交叉验证的方法来选择最佳参数, 这个过程被重复了三次。图 2-7 展示建立的五个模型, 分别是: (a) 仅使用气候数据, (b) 使用气候和土壤数据, (c) 使用气候、土壤和 NDVI 数据, (d) 使用气候、土壤和 EVI 数据, 以及 (e) 使用气候、土壤和 NIRv 数据。其中表中 M1: 仅气候数据; M2: 气候+土壤数据; M3: 气候+土壤+NDVI 数据; M4: 气候+土壤+EVI 数据; M5: 气候+土壤+NIRv 数据。

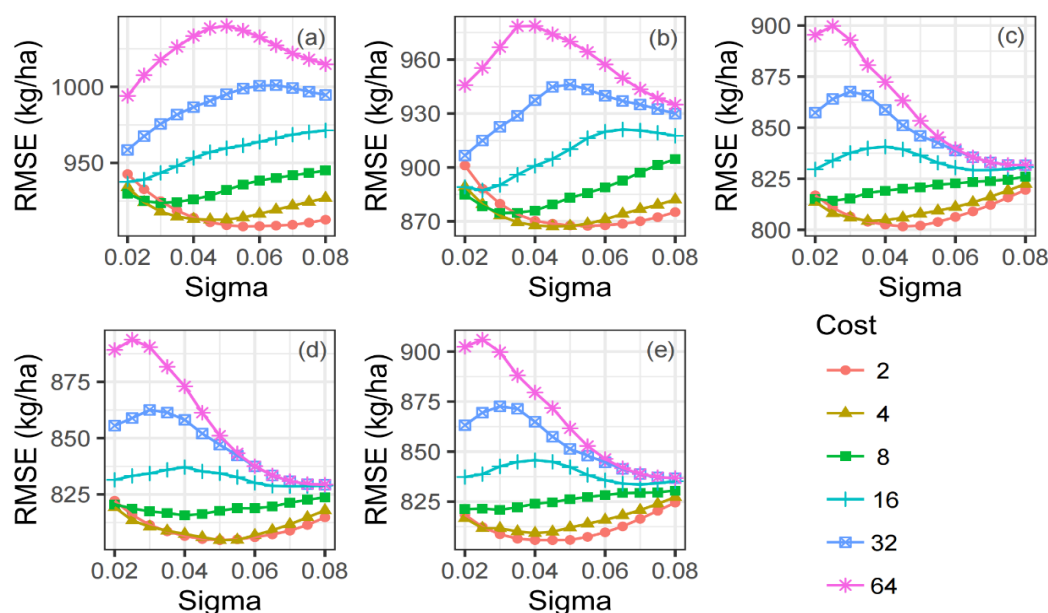


图 2-7 基于训练数据计算的支持向量机模型的不同参数的 RMSE

Fig. 2-7 The cross-validation error rates (RMSE) in tuning parameters (Cost and Sigma) for SVM using training data sets for predicting wheat yield

### 2.1.3.3 随机森林

随机森林模型 (RF) 是一种基于分类和回归树集合的非参数方法 (Breiman 2001)。每棵树都是抽样构建的，留下大约三分之一的样本进行验证。每棵树都返回均值或平均预测值，以提高该数据集的性能。RF 模型可以捕获产量和预测变量之间的关系 (非线性或线性)。最近，RF 方法被广泛用于产量估算，以及确定产量对气候因素的响应 (Hoffman et al. 2020)。在参数选择中把  $mtry$  和  $ntree$  分别设置为 1-25 (间隔为 2) 100-900 (间隔为 200)。通过这样的方式为每次建模提供了最优的参数组合 (图 2-8)。图中的 (a) - (e) 分别代表建立的五个模型：(a) 仅气候数据，(b) 气候+土壤数据，(c) 气候+土壤+NDVI 数据，(d) 气候+土壤+EVI 数据，(e) 气候+土壤+NIRv 数据。最终的参数见表 2-1。

RF 算法还可以评估每个预测变量的重要性。本研究首先汇总了每个站点四个生长期 (T1-T4) 的多源环境变量以进行产量预测。然后在 RF 模型中应用这些预测因子来预测作物产量。每个变量的相对重要性由“%IncMSE”指标评估。将相对重要性高的变量确定为影响小麦产量的主要因素，使用 RF 模型的部分依赖图来估计作物产量对生长期预测因子的响应。

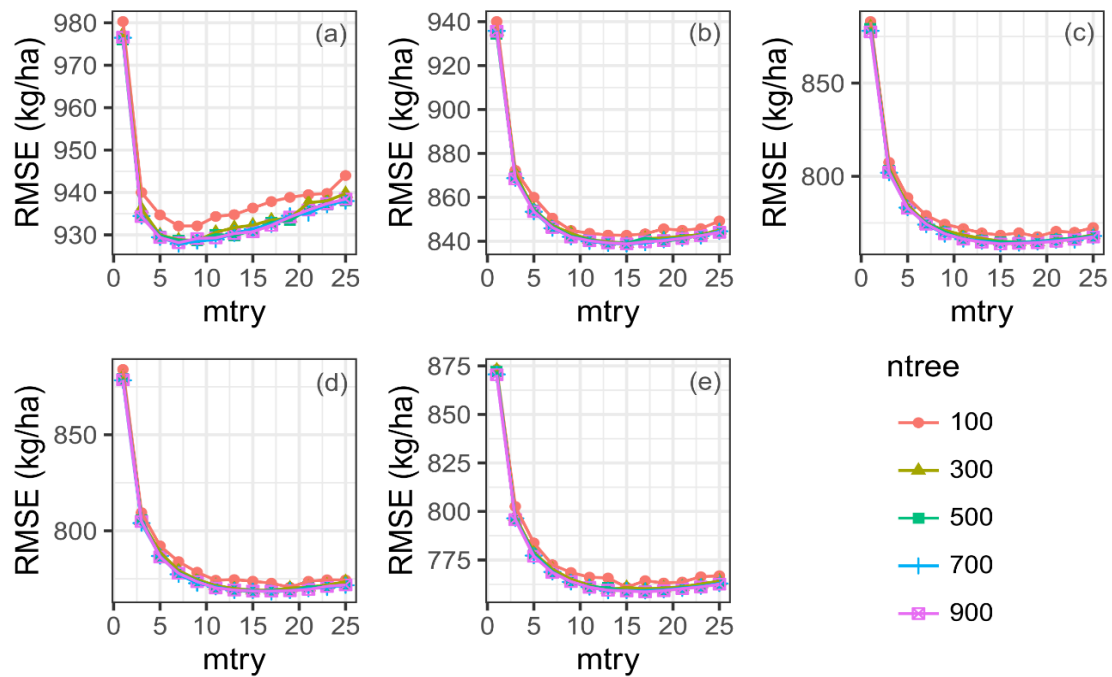


图 2-8 基于训练数据计算的随机森林模型的不同参数的 RMSE

Fig. 2-8 The cross-validation error rates (RMSE) in tuning parameters (mtry and ntree) for RF using training data sets for predicting wheat yield

表 2-1 小麦产量预测中不同输入数据集下 RF 和 SVM 模型的优化参数

Table 2-1 Optimal parameters for RF and SVM model with different input data sets in wheat yield prediction

机器学习方法	RF		SVM	
参数	ntree	mtry	Cost	Sigma
M1	700	7	2	0.055
M2	700	15	4	0.045
M3	900	15	2	0.045
M4	700	17	2	0.05
M5	900	17	2	0.045

RF 和 SVM 都重复运行 100 次。每次运行都使用随机选择的总数据的 70% 的值进行训练，并使用总数据的 30% 对整个研究区域进行验证。本研究使用决定系数 ( $R^2$ ) 和均方根误差 (RMSE) 来评估 RF 和 SVM 模型的性能。

## 2.2 结果分析

### 2.2.1 模型性能评价

对每个模型进行了 100 次重复运行，并对性能进行了评估。结果表明，在 SVM 和 RF 模型中，使用 M3、M4 和 M5 的输入数据比 M1 和 M2 具有更好的性能（见图 2-9）。这表明使用更多信息量的预测变量会导致更高的产量估计精度。使用 M5 数据产生的性能略好于 M3 和 M4，最高平均  $R^2$  值（RF, 0.74; SVM, 0.69）和最低平均 RMSE 值（RF, 758 kg/ha; SVM, 821 kg/ha）。这些结果表明，与使用 EVI 和 NDVI 相比，包括 NIR<sub>v</sub> 可以略微提高小麦产量预测的性能。此外，无论使用何种数据集，RF 在小麦产量预测方面都比 SVM 表现出更好的性能（见图 2-9）。误差棒表示 100 次运行的标准差。RF 表示随机森林，SVM 表示支持向量机，M1 至 M5 表示不同的数据集。

使用具有气候+土壤+ NIR<sub>v</sub> 数据的模型作为最终模型来评估三个子区域的模型性能（图 2-10）。结果表明模型的性能在三个子区域之间存在差异。总的来说，小麦预测准确率最高的是分区 III ( $R^2 = 0.82$ , RMSE = 685 kg/ha)，分区 II 次之 ( $R^2 = 0.66$ , RMSE = 761 kg/ha)，分区 I 次之。区域 I ( $R^2 = 0.1$ , RMSE = 890 kg/ha)。RF 模型在子区域 I 表现不佳主要原因可能由于该区域水分主要依靠灌溉，然而由于数据量不足和一些农艺措施（如灌溉）没有被多源环境数据反映。RF 模型在子区域 II 和 III 中表现较好，SVM 模型在子区域 I 中表现稍好。

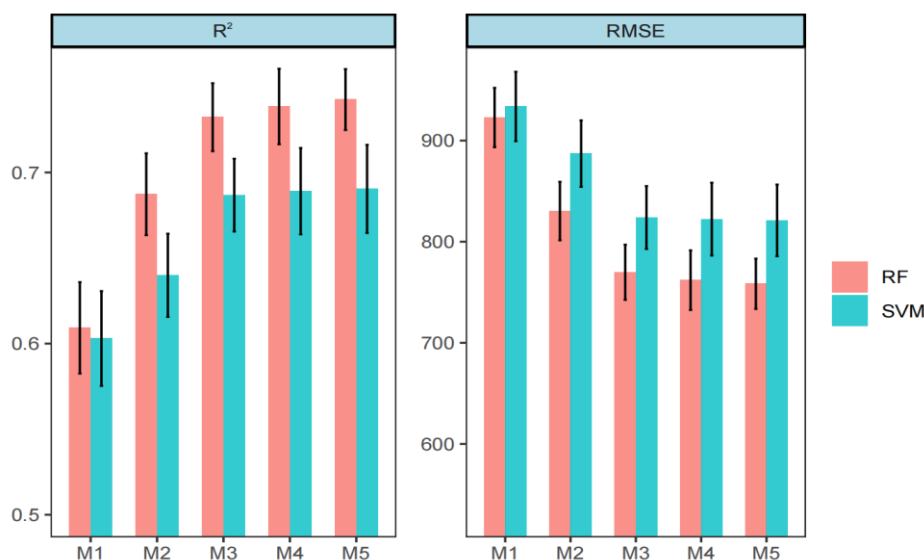


图 2-9 基于多源环境变量的模型精度评价，每个模型基于 2001 年至 2020 年期间的 100 次重复运行进行评估。

Fig. 2-9 Model performance of wheat yield prediction using different multi-source data. Each model was evaluated based on 100 repeated runs during 2001–2020.

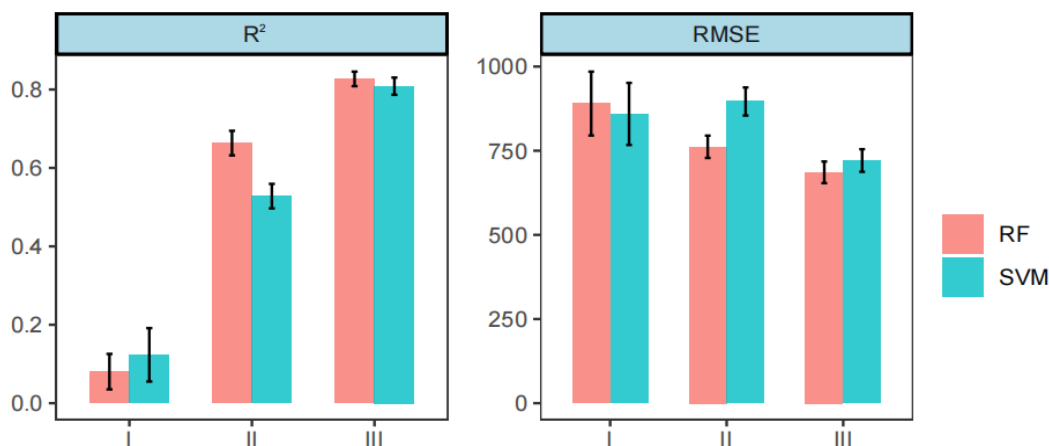


图 2-10 RF 和 SVM 模型使用气候+土壤+NIR<sub>v</sub> 数据在不同子区域的表现。RF:随机森林;SVM:支持向量机;M1:仅气候数据;M2:气候+土壤数据;M3:气候+土壤+ NDVI 数据;M4:气候+土壤+ EVI 数据;M5:气候+土壤+ NIR<sub>v</sub> 数据

Fig. 2-10 Testing RF and SVM with climate + soil + NIR<sub>v</sub> data in different sub-regions.RF: Random Forest; SVM: Support Vector Machine; M1: climate data only; M2: climate + soil data; M3: climate + soil + NDVI data; M4: climate + soil + EVI data; M5: climate + soil + NIR<sub>v</sub> data.

## 2.2.2 变量的相对重要性

模型 RF\_M3、RF\_M4 和 RF\_M5 中使用的预测变量的相对重要性如图 2-11 所示。对于植被指数，VI 的重要性对于基于 NDVI 的模型（RF\_M3）为 27%，对于基于 EVI 的模型（RF\_M4）为 29%，对于基于 NIR<sub>v</sub> 的模型（RF\_M5）为 30%。对于所有三个模型，土壤特性的重要性是相同的（10%）。所有与气候相关的预测因子的相对重要性约为 60%（RF\_M3：63%、RF\_M4：61% 和 RF\_M5：60%）。

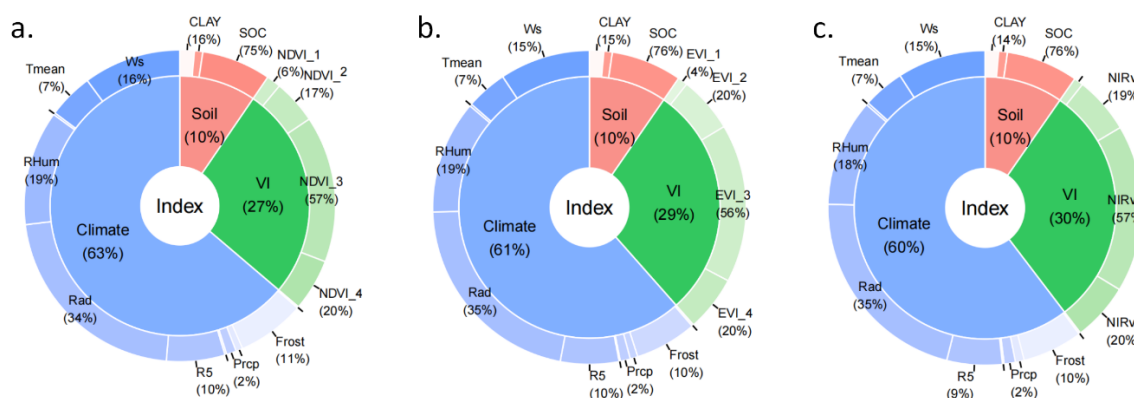


图 2-11 不同多源环境变量（气候、土壤和植被指数）的相对重要性比例。

Fig. 2-11 Proportion of relative importance of different sources of environmental data (climate, soil, and vegetation indices)

图 2-12 显示了每个模型的前 15 个变量的相对重要性。结果一致表明，在三个模



型中,小麦产量主要受 T3 期间的 VI(NDVI\_3、EVI\_3 和 NIRv\_3)以及 T2(Rad\_2)和 T3(Rad\_3)期间的辐射影响(图 2-12)。

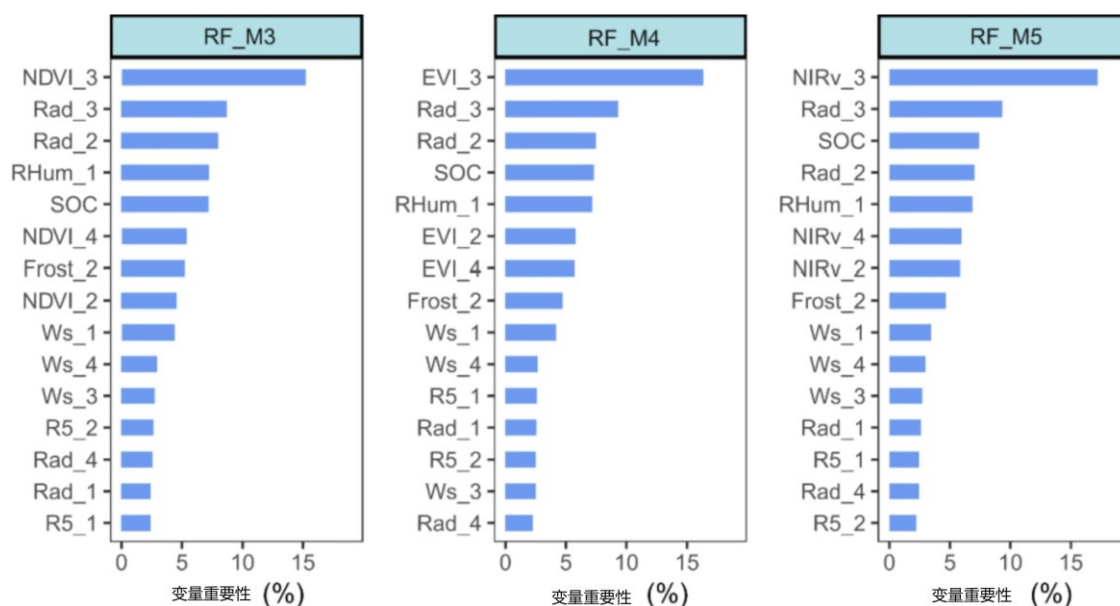


图 2-12 随机森林模型变量重要性排名前 15 的变量

Fig. 2-12 Relative importance of predictor variables (ranked for the first fifteen) from different RF models

### 2.2.3 小麦产量对预报因子的响应

本研究基于RF\_M5模型进一步研究了小麦三个生育期的NIRv如何影响产量。NIRv的局部依赖图(PDP)揭示了小麦产量与NIRv\_2(19.1%)、NIRv\_3(57.4%)和NIRv\_4(19.9%)之间的关系(图2-13)。麦产量与NIRv的响应曲线呈近线性关系。例如,当NIRv\_2在0.01~0.16范围内时,小麦产量会随着NIRv\_2的增加而大幅增加。然而,当NIRv\_2大于0.16时,小麦产量趋于稳定。当NIRv\_3在0.02-0.21范围内时,随着NIRv\_3的增加,小麦产量将大幅增加。然而,当NIRv\_3超过0.21时,小麦产量趋于稳定。NIRv\_4也有类似的结果。图中蓝色阴影区域代表第10和第90的百分位数之间的校准数据。百分比值表示从随机森林模型生成的每个预测变量的相对重要性。箱形图表示不同子区域NIRv值的可变性和范围。

土壤特性数据也极大地促进了产量估算。然而,这些数据不是动态数据(它们不随时间变化)。因此,本研究没有显示土壤特性的PDP。气候因素强烈影响不同生育期的小麦产量。气候预测因子(排名前十二的气候预测因子)的PDP代表了作物产量对每个变量的响应曲线,如图2-14。每个次区域的气候条件各不相同(图2-14所示的箱线图),表明气候相关因素的影响在不同的子区域可能有所不同。例如,T3期间的辐射量(Rad\_3)是产量预测的主要影响因素,当Rad\_3在16 MJ/m<sup>2</sup>左右时预测小麦

产量最高（图2-14a）。值得注意的是，子区域III中的辐射较低( $\text{Rad}_3 < 16 \text{ MJ/m}^2$ )可能会限制小麦产量。在 T2( $\text{Rad}_2$ )期间，也发现了类似的规律（图2-14b）。作物产量也受到RHum\_1的影响(当  $\text{RHum}_1 > 61\%$ )（图2-14c）。当  $\text{Frost}_2 < 90$  天时， $\text{Frost}_2$  对小麦产量有积极影响。当 $\text{Frost}_2 > 90$ 天时，小麦产量略有下降（图2-14d）。风速在各个生育期与小麦产量密切相关，当风速低于 5-6 m/s 时表现出正向影响(图2-14e-g)。小麦产量在 T1（图2-14 h）和 T4（图2-14 j）期间对 Rad 表现出相似的反应。R5\_1（图2-14i）、R5\_2（图2-14 k）和 Tmean\_2（当  $\text{Tmean}_2 > 4^\circ\text{C}$ 时）（图2-14l）显示出对小麦产量的负面影响。图2-14中蓝色阴影区域代表第10和第90百分位数之间的数据。百分比值表示从RF\_M5模型生成的每个预测变量的相对重要性。箱线图显示了不同子区域中每个气候变量的变异性范围和范围。

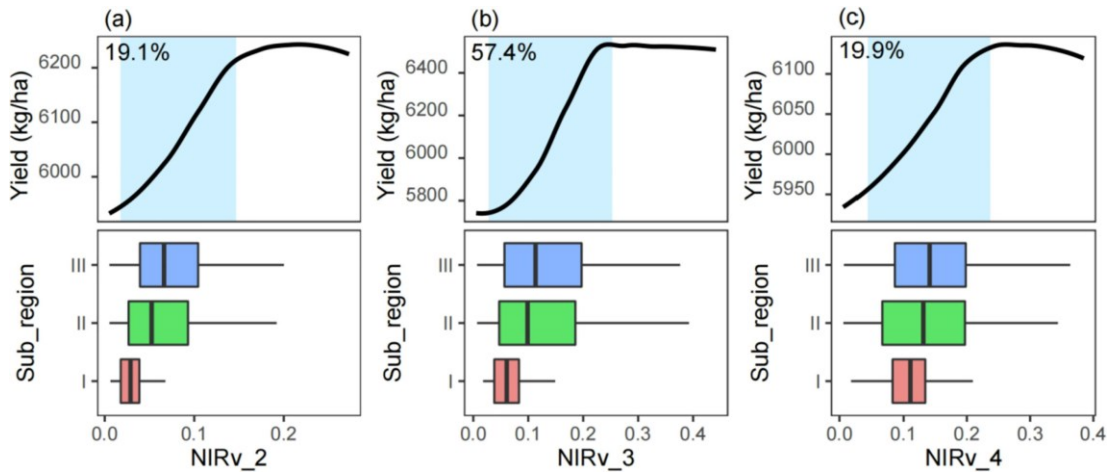


图 2-13 基于 RF\_M5 模型的四个生长阶段相对重要性值排名前三的 NIRv 与产量的局部依赖图。黑线是经过平滑处理的拟合值。这条线反应的是趋势，而不是实际值，描述了小麦产量依赖于预测因子的性质。蓝色阴影区域表示第 10 分位数和 90 分位之间的值。百分比值表示由随机森林模型生成的每个变量的相对重要性。箱形图显示了不同子区域 NIRv 值的变化范围。

Fig. 2-13 Partial dependence plot of NIRv with the relative importance value ranked in the first three during four growth stages based on the RF\_M5 model. The black lines are smoothed representations of the response with fitted values (model predictions) for the calibration data. The trend of the line, rather than the actual values, describes the nature of the dependence of wheat yield on the predictors. The blue shaded area represents calibration data between the 10th and 90th percentile. The percentage values represent the relative importance of each predictor generated from the random forest model. The box plots indicate the variability and range of NIRv values in different sub-regions.

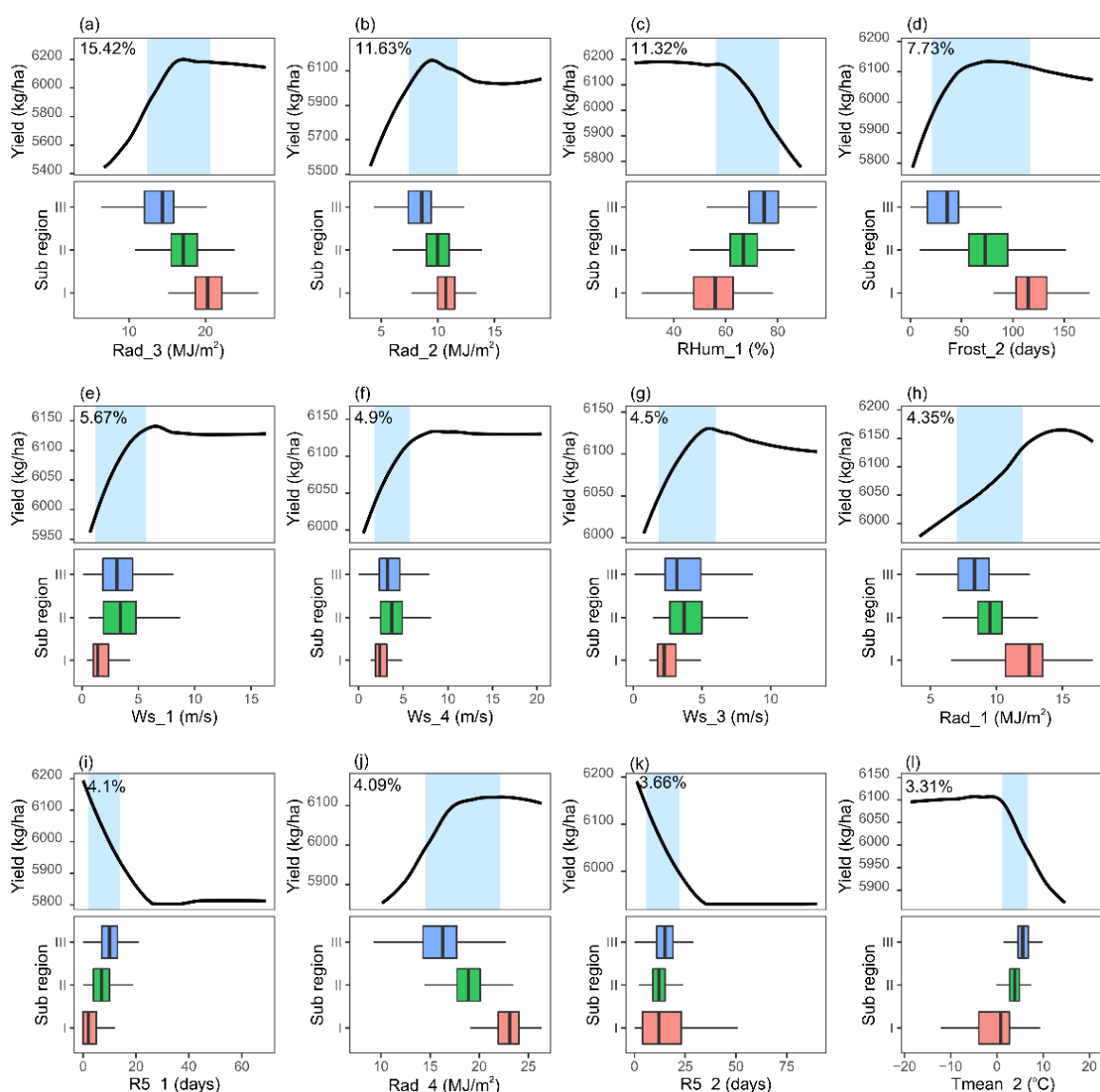


图 2-14 使用 RF\_M5 中气候预测因子中不同气候变量重要性排前十二的部分依赖图。黑线是经过平滑处理的拟合值。这条线反应的是趋势，而不是实际值，描述了小麦产量依赖于预测因子的性质。蓝色阴影区域表示第 10 分位数和 90 分位之间的值。百分比值表示由随机森林模型生成的每个变量的相对重要性。箱形图显示了不同子区域气候要素的变化范围。

Fig. 2-14 Partial dependence plots of different climate variables with feature importance ranked for the first twelve climate predictors in different sub-regions using model RF\_M5. The black lines are smoothed representations of the response, with fitted values (model predictions) for the calibration data. The trend of the line, rather than the actual values, describes the nature of the dependence of wheat yield on the predictors. The blue shaded area represents calibration data between the 10th and 90th percentile. The percentage values represent the relative importance of each predictor generated from the RF\_M5 model. The box plots indicate the variability and range of each climate variable in different sub-regions.

其它气候因子的PDP如图2-15所示：

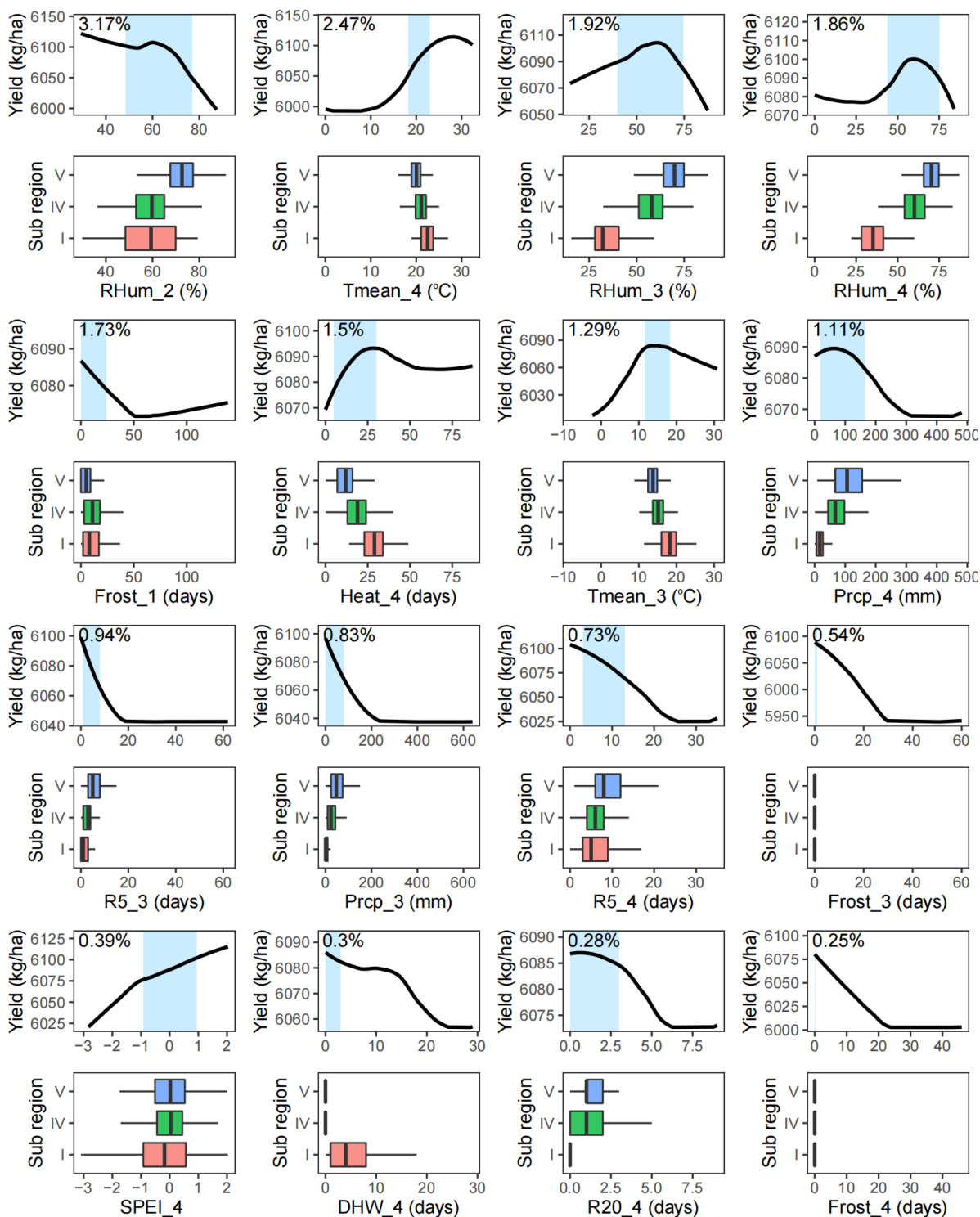


图 2-15 使用 RF\_M5 中气候预测因子中排名 12 之后的依赖图。

Fig. 2-15 Partial dependence plots of different climate variables with feature importance ranked after twelve climate predictors in different sub-regions using model RF\_M5.

### 2.2.4 不同子区域小麦的主要影响因子

不同子区域和生育时段的主要气候因子也会有所不同。本研究进一步分析了不同子区域产量的主要限制因子（图 2-16）。

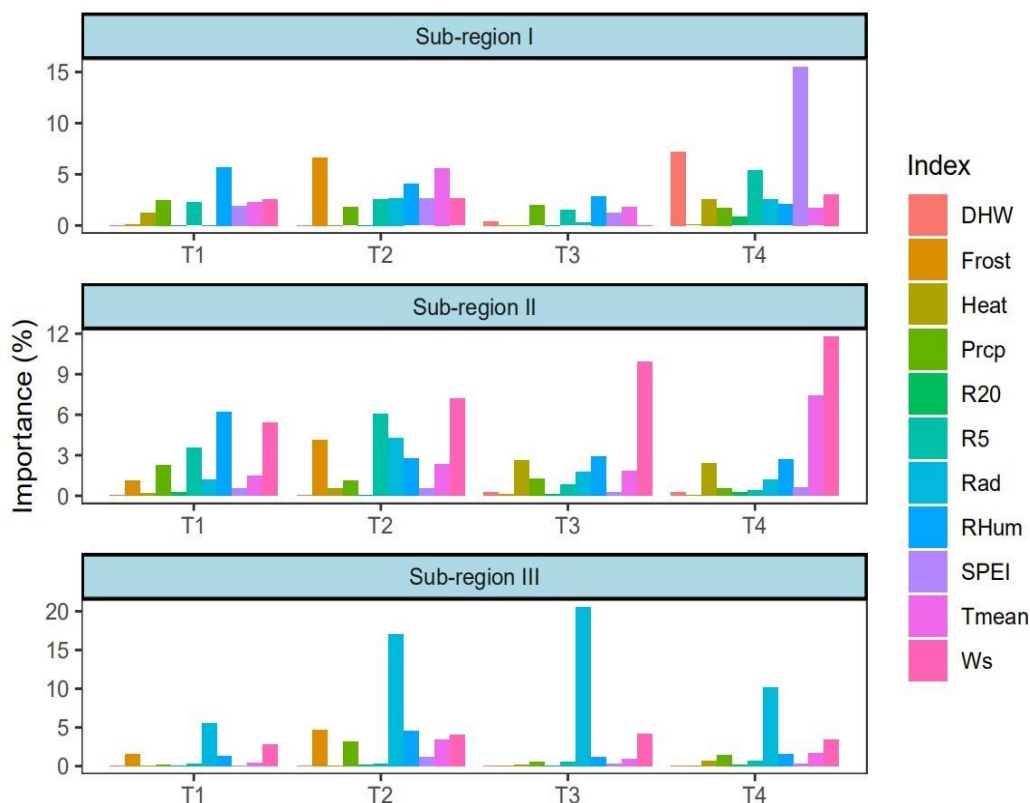


图 2-16 不同分区四个生育时段气候因子重要性。

Fig. 2-16 The relative importance of climate factors in different sub-region during four growth stages.

图 2-16 的结果表明，在西北荒漠区（I）干旱是产量的主要限制因子，尤其是在 T4 时段，说明在产量形成期干旱对该区域有较大的影响。主要原因是该区域属于典型的干旱区，导致小麦生长受到较大的限制（Yao et al. 2020）。干旱对冬小麦产量的影响主要是通过水分限制作用于其生长和发育的关键时期，特别是生育后期，导致小麦形成不健康的生育状态，从而降低了小麦的产量（Chen et al. 2020d）。因此，建议该区域干旱方面的管理仍然需要加强。例如通过加强水土保持工程，来提高土壤保水能力；改善灌溉技术，减少水分浪费；培育新的抗旱品种等。对于华北平原来说（II），风速对小麦有较大影响。这可能是由于风速会加速水分蒸发并导致病虫害风险加剧，因此对产量会有一定的影响。相反，适宜的风速可以降低作物叶片表面的温度和湿度，有助于增强叶片的光合作用能力，从而促进作物的生长和发育。在区域 III 中，辐射是主要的限制因子，这可能是由于该区域多数区域处于四川盆地，冬季太阳辐射强度低，日照时间短，加之云雾频繁，会导致冬小麦受到热量不足的影响，影响其生长和



发育，从而影响产量。

为了进一步探讨这些因子在不同区域与小麦产量的非线性关系，本研究分析了重要性靠前的因子与产量的局部依赖图（图 2-17）。图中，对于区域 I，可以看到干热风 and 干旱（SPEI）复合会加剧作物减产。对于区域 II，T4 阶段，温度在 19-23°C 产量最高。T4 阶段风速随产量增加而增加，但是在 T2 阶段，当风速大于 7m/s 时产量将有下降。辐射在区域 III 对产量有较大的促进作用。此外，空气相对湿度在 62-75% 时产量最高。

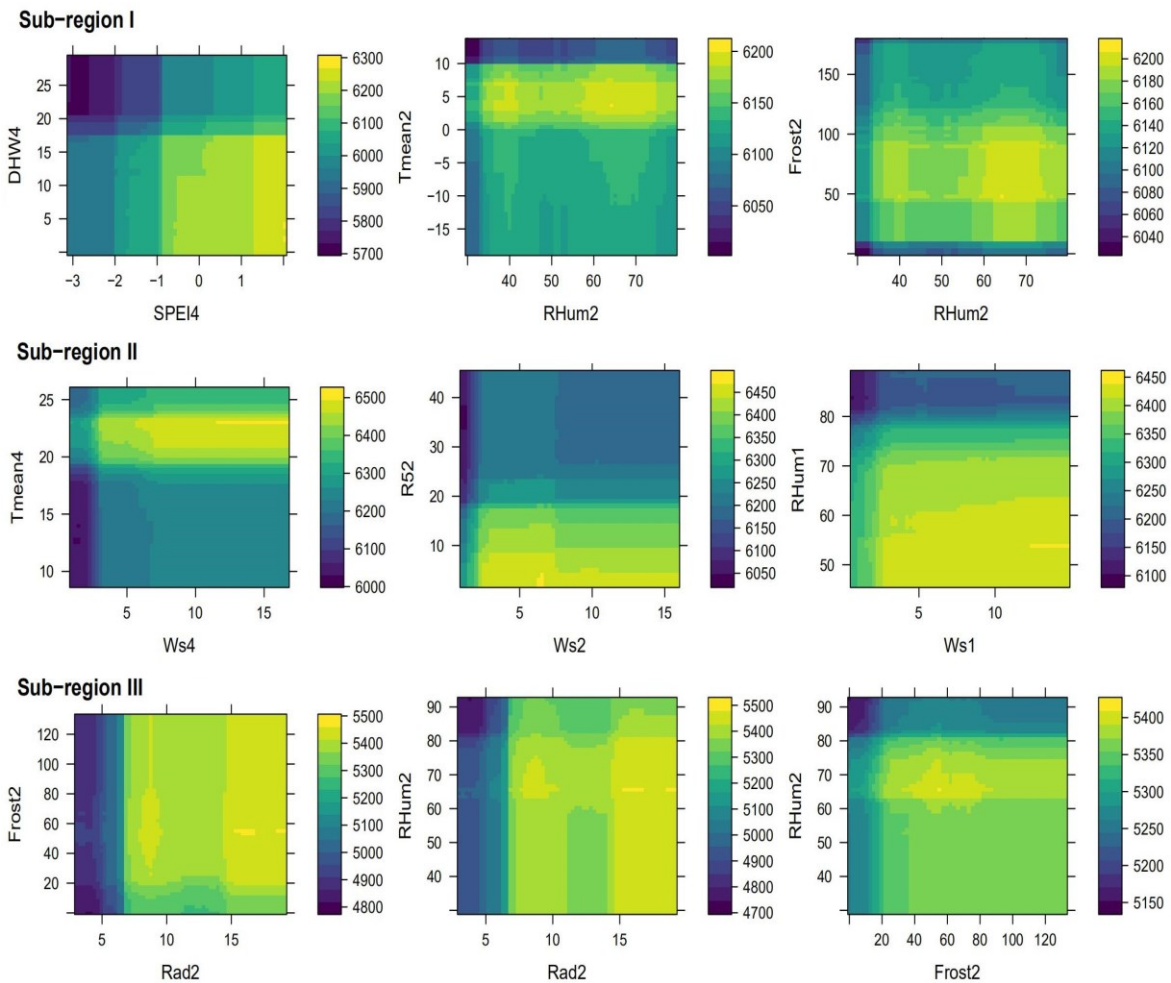


图 2-17 三个不同分区主要影响因子的局部依赖关系图。

Fig. 2-17 Partial dependence plot of the main influencing factors in the three sub-regions.

## 2.3 讨论

### 2.3.1 模拟精度评价

本研究的结果表明，基于多源环境数据的机器学习模型可以在田间规模上提供可靠的小麦产量预测。先前的研究预测了县级的小麦产量。例如，Wang 等（2020g）用

卷积神经网络(CNN) 和长短期记忆 (LSTM )预测中国的冬小麦产量, 并显示出可接受的精度 ( $R^2$  为 0.74, RMSE 为 721 kg/ha)。同样 Han 等 (2020) 使用 RF、SVM 和高斯过程回归 (GPR) 预测县级产量,  $R^2$  高于 0.75。Cao 等 (2021b) 使用随机森林和三种深度学习算法 (CNN、LSTM 和深度神经网络) 结合多源环境变量预测了县级小麦产量并得到了比较好的精度 ( $R^2 \geq 0.85$ ,  $RMSE \leq 768 \text{kg/ha}$ )。同时, 他们也做了站点尺度的预测, 其中在站点尺度上  $R^2=0.71$ 。值得注意的是在田间预测作物产量更加困难, 因为即使在同一个县, 环境条件也可能变化很大, 因此需要更高分辨率的数据集。与过去站点尺度预测的结果相比 ( $R^2=0.71$ ) (Cao et al. 2021b), 本研究的精度有了略微提高 ( $R^2=0.74$ ), 这可能是因为本研究结合了更长的事件序列(2001-2020) 和更加庞大的数据集, 以及考虑了一些新的植被指数。

### 2.3.2 植被指数对产量预测的贡献

进行了多组数据集和机器学习方法的模拟精度比较, 结果显示基于 NIRv 的随机森林模型优于基于 EVI 和 NDVI 的模型。这表明 NIRv 能够更好地捕捉作物生长和产量形成的信息, 因为它能够有效地分离土壤背景与植被信号, 同时能够区分光合作用与冠层深度的分布 (Badgley et al. 2017; Huang et al. 2019; Ryu et al. 2019)。此外, 本研究还发现相比于其他作物, NIRv 能够更好地捕捉到产量的胁迫, 而且与 GPP 有很好的相关关系 (详见图 2-18)。Peng 等 (2020a) 评估了不同基于卫星的植被指数 (例如, NDVI、EVI、地表温度、NIRv 和 SIF ) 和 SIF 数据 (OCO-2、GOME-2 和 TROPOMI) 用于预测美国玉米和大豆产量的能力。他们也发现 NIRv 用于预测美国玉米和大豆产量的能力比其他植被指数更为精准, 甚至可以与 SIF 产品 (OCO-2 和 TROPOMI) 媲美或者更好。这可能是由于相比于 NDVI, NIRv 与 GPP 有强烈的相关性, 且能够更早地检测到气候胁迫, 如温度胁迫 (Badgley et al. 2019; Wang et al. 2020f)。NIRv 是一种基于 MODIS 的指数, 具有比 SIF 产品更高的空间分辨率 ( $<1 \text{ km}$ )。尽管 NIRv 是基于 MODIS 的指数, 但其空间分辨率高于 SIF 产品 ( $<1 \text{ km}$ )。然而, 最近出现的高时空分辨率的 SIF 产品, 如 GeoCarb 和 TEMPO, 为基于 SIF 的作物产量预测提供了巨大的潜力。

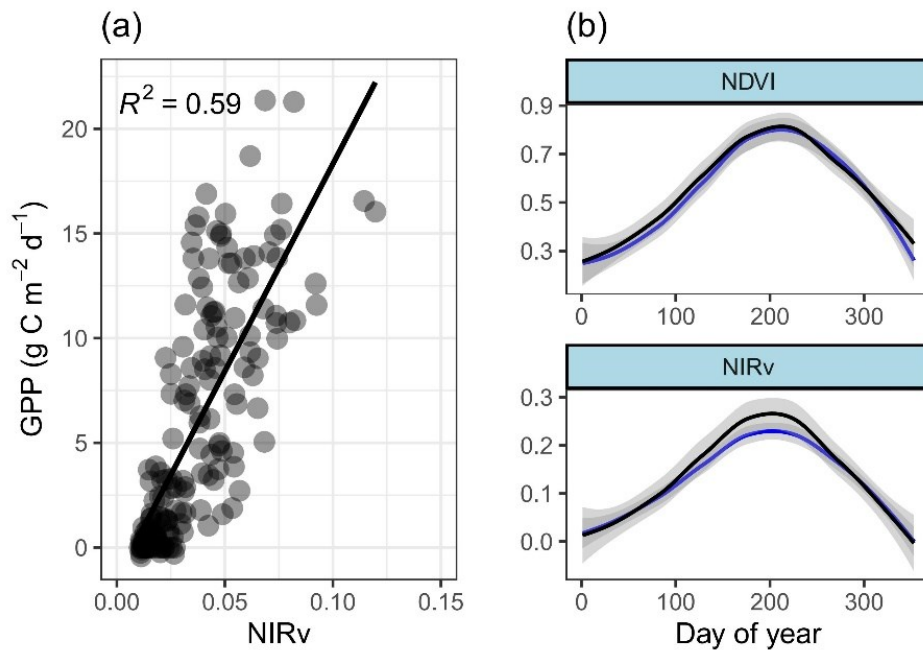


图 2-18 NIR<sub>v</sub> 的评价。(a) 2003-2010 年禹城站 GPP 与 NIR<sub>v</sub> 的相关性。NIR<sub>v</sub> 与 GPP 有较强的相关性。(b) 为 2011 年湖北省干旱事件，平滑的黑蓝线代表 NIR<sub>v</sub> 和 NDVI 的时间序列变化。与非干旱年份相比，NDVI 在夏季峰值上差异不大，而整个夏季的 NIR<sub>v</sub> 都表现出抑制作用。

Fig. 2-18 The performance of NIR<sub>v</sub>. (a) The correlation of GPP and NIR<sub>v</sub> in Yucheng station during 2003-2010. NIR<sub>v</sub> showed a strong correlation with GPP. (b) Showed the drought event during 2011 in Hubei province, the black and blue lines are smoothed represented the time-series change of NIR<sub>v</sub> and NDVI. The NIR<sub>v</sub> showed suppression throughout the summer months when compared against the non-drought year, while NDVI showed little difference in peak summer values.

结果表明 T3 期间的 NIR<sub>v</sub> 是影响小麦产量估计的最重要预测因子（图 2-12）。这主要是因为分蘖数和叶面积量在此时间间隔内达到峰值（Shao et al., 2013）。因此，可以通过植被指数检测这一时期的作物生长状况和光合作用。结果表明，当 NIR<sub>v</sub> 低于 0.21 的阈值时，预测的小麦产量与 NIR<sub>v\_3</sub> 几乎呈线性相关，表明光合速率是小麦产量的主要限制因素。然而，一旦 NIR<sub>v</sub> 超过 0.21，小麦产量几乎没有变化（图 2-13）。表明其他因素对小麦产量和光合作用的潜在不对称影响，例如光合作用效率（He et al. 2020; Liu et al. 2017c）、温度或太阳辐射因素（Chen et al. 2020a）。此外，由于作物生长、发育和环境条件相互作用的反馈影响，光合作用并不总是与最终作物产量直接相关（Wu et al. 2019）。因此，推测当 NIR<sub>v</sub> 大于该阈值时，作物产量会受到其他因素（如气候条件和作物氮素状况）的限制。由于 NIR<sub>v</sub> 可以作为反映光合的替代指标，且光合作用与产量关系具有很大的相关关系。分析了 NIR<sub>v</sub> 与气候因子的响应关系，并对比了其非线性关系的阈值，发现整体上光合响应曲线与产量与气候的响应曲线比较接近（图 2-19）。说明气候因子对农作物产量的影响主要是通过影响光



合作用来实现的。同时，这也表明了植物的光合作用是决定农作物产量的重要因素之一。

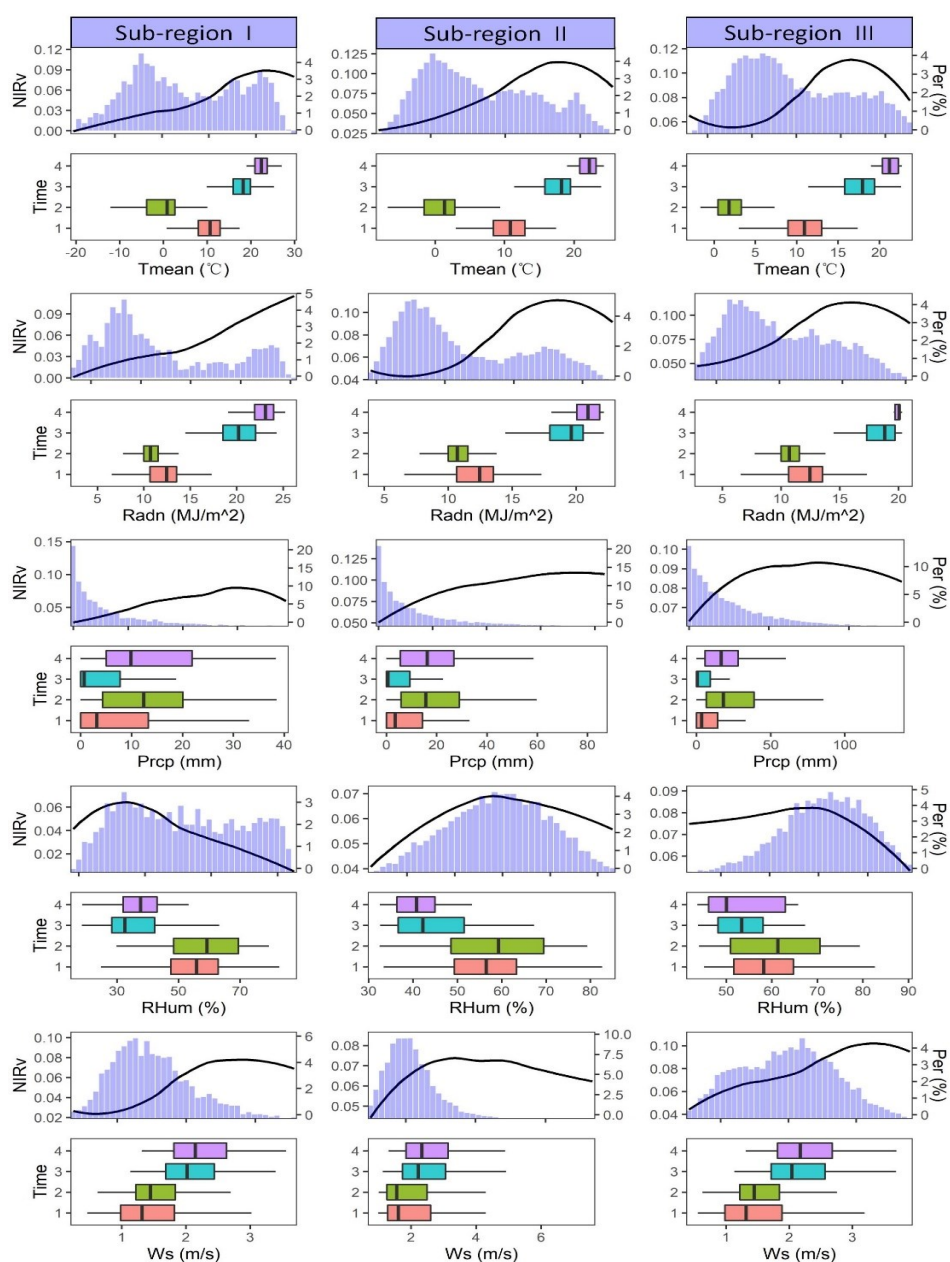


图 2-19 2001-2020 年不同分区气候与 VIRv 响应曲线是 NIRv 中值(左 y 轴)对不同气候变量的响应的平滑表示。直方图(右 y 轴)显示了 2001-2020 年各气候变量的百分比。箱形图表示各气候变量在不同时间间隔(T1-T4)的变化范围

Fig. 2-19 The response curve of climate and VIRv in different sub-regions during 2001-2020. Black lines are smoothed representations of the median NIRv value (left y axes) respond to different climate variables. Histograms (right y axes) showed the percentage of each climate variable in 2001-2020. The box plots indicate the range of each climate variable at different time intervals (T1-T4).

### 2.3.3 气候因素与产量的关系

研究的结果表明,太阳辐射在 T3 和 T2 期间是除 NIRv\_3 之外最重要的气候变量(见图 2-12)。由于辐射强度与光合作用之间存在密切联系,而 NIRv 可作为光合作用的替代指标,因此辐射对产量的响应曲线与产量和 NIRv 的响应曲线非常相似(见图 2-13 和图 2-14)。此外,管理措施(例如灌溉)以及其他气候相关因素(如干旱)可能会减弱它们对产量的影响。当 RHum\_1 > 61%时,T1 期间的相对湿度也对小麦产量产生负面影响(见图 2-14)。这可能是由于高相对湿度会增加作物病虫害的风险。值得一提的是,SOC 在 RF\_M5 模型中也非常重要,因为 SOC 具有许多附加效益,如提高土壤质量和作物生产力,正如先前的研究所展示的(Hammad et al. 2020; Lal 2006; Majumder et al. 2008)。当 Frost\_2 < 90 天时,Frost\_2 对小麦产量有积极影响。这是因为小麦需要较低的温度才能完成春化,并且在营养早期具有较高的抗冻性(Bergjord et al. 2008; Xiao et al. 2018)。此外,研究的结果表明,T1、T4 和 T3 期间的微风会提高小麦的产量(见图 2-14),而小雨则会降低小麦的产量。尽管破坏性暴雨可能会直接破坏农田,但从国家层面来看,这种可能性相对较低。相对于小雨而言,下小雨的时候通常伴随着低辐射和低温,且小雨提供的水分对作物的生长效果有限。此外,下小雨的时候空气湿度较高,为真菌病原体提供了适宜的条件,也会增加病虫害的风险(Harvell et al. 2002; Lesk et al. 2020; Sun and Woods 1994)。在较低风速时,对产量也有一定的促进作用。例如,叶片响应微风而颤动可能导致太阳辐射更好地渗透到较低的冠层,从而提供更多机会来维持更大的冠层光合作用(Burgess et al. 2019; Roden and Percy 1993)。

在不同的子区域中,由于气候条件的不同,气候相关因素的影响也各不相同。例如,太阳辐射是影响小麦产量的主要因素之一,而由于 III 区(如四川盆地)的太阳辐射低于 I 区和 II 区,因此应更加关注其影响。此外,霜冻在分区 I 和分区 III 表现出相反的作用(见图 2-14)。这是由于 I 分区霜冻事件频繁发生,霜冻日数高于阈值,导致小麦产量受损较大。因此,在气候相关影响研究中,应考虑特定区域的特点。

总体来看,本研究量化了小麦产量对气候变量和极端气候的反应,为减少气候造成的产量损失风险提供了有价值的信息。研究发现,未来干旱、高温和极端降水可能更加频繁发生,这将导致气候条件不利于作物生长(Li et al. 2019a; Li et al. 2019b; Yao et al. 2020)。一些研究通过使用基于过程的作物模型,评估了几种对气候变化的适应性(Challinor et al. 2014),例如培育新品种、优化播种日期、使用灌溉和管理作物残留物。但是,作物模型对于极端气候事件过于简化,可能会低估一些负面影响。因此,应该寻求通过将机器学习与生物物理模型相结合来开发混合模型,以提供更为稳健的气候变化影响评估,并探索不同适应方案的潜力。

### 2.3.4 研究的局限性

本章的工作有一些局限性，首先，机器学习是数据驱动模型。模型的性能在很大程度上取决于数据的质量以及数量。这可能会直接影响研究的结果。第二，一些开发出来的基于高分辨率 SIF 产品（例如，OCO-2 和 TROPOMI），可能可以提高模型的模拟精度（Zhang et al. 2018）。最近，Camps-Valls 等（2021）提出了一个新的植被指数，kNDVI，它与 GPP 的相关性强于 NDVI 和 EVI。此外，Dechant 等（2020b）的研究表明 NIRvP（计算为  $\text{NIRv} \times \text{PAR}$ ）是远红外 SIF 的有力代表，并且与 SIF 的相关性强于其他指数。这些指标可能对提高作物生物量预测有很大的潜力，本研究未考虑这些最新的植被指数。第三，本研究虽然可以较准确的模拟产量，但是却无法提前预报产量，如果能提前 1-3 月较为可靠的预报出产量，这样可以为决策者留下更多的时间来及时的调整措施，并制定相应的策略。

## 2.4 小结

本研究开发了基于机器学习与多源环境变量（包括土壤特性、气候和植被指数）的模型来预测小麦产量。得出以下结论：

（1）RF 在预测小麦产量方面优于 SVM。使用 NIRv ( $R^2=0.74$ ;  $\text{RMSE}=758$  kg/ha) 的 RF 模型比使用 EVI ( $R^2=0.73$ ;  $\text{RMSE}=762$  kg/ha) 或 NDVI ( $R^2=0.73$ ;  $\text{RMSE}=770$  kg/ha)。此外，与其他环境协变量相比，基于植被的指数对小麦产量的影响最大。

（2）基于 RF\_M5，发现 T3 期间的 NIRv 是决定作物产量的最重要预测因子。此外，Rad (T2 和 T3)、RHum (T1)、SOC、Ws (T1-T4)、R5 (T1-T2) 和 Frost (T2) 被确定为限制小麦产量的主要因素。干旱对产量变化的贡献相对较小，因为在本研究中好多小麦处于灌溉区域。

（3）构建了 PDP 图来解释每个发育时期的不同预测因子如何基于 RF 模型影响小麦产量。发现 NIRv 与小麦产量呈线性和非线性关系。小麦产量对其他环境变量有类似阈值的反应。这些 PDP 结果有助于更好地了解这些因素如何限制小麦产量。

研究结果证明使用 NIRv 进行产量预测可以得到相对较好的精度，并且本研究的建模方法使用公开数据广泛适用于全球其他地区。产量预测模型可以在未来通过结合基于过程的作物模型或其他新开发的植被指数得到进一步改进。

### 第三章 建立动态产量早期预报模型并识别关键生长阶段

在预测作物产量方面，大多数研究都是在作物生长期末阶段进行模拟，而不是在收获前进行预测。本研究基于上一章的结果，采用收获前的数据建立模型，能够更早地预测作物产量，并确定最优预报时段和关键生长期。这样的早期可靠预测信息对于农民和决策者都非常重要，因为它可以为利益相关者提供在各自工作中做出决策所需的重要支持。一旦特定地点的农作物产量预测完成，就可以确定适当的农场管理实践和安全预防措施，例如粮食储存等。产量预测还可以为政府决策者和粮食营销机构提供宝贵信息，帮助规范农产品市场和制定贸易策略。然而，目前对于不同粮食作物构建产量预测模型的研究较少，仅有的一些研究都是基于单一作物来构建模型，而政策制定者更希望分析和对比不同作物，以更好地制定相关政策。因此，本章针对不同作物（包括冬小麦、春玉米、夏玉米、早稻、中稻和晚稻）基于多源环境数据构建了产量预报模型，并分析了决定作物产量的关键预报因子。具体研究目标包括：（1）利用我国主要粮食产区的多源环境数据，基于随机森林模型创建多个模型来预测三种主要作物的产量；（2）确定不同作物产量预测精度可接受的最佳提前期以及关键的生长阶段。本研究对于提高中国农业生产效率和质量，以及农业市场的规范化和可持续发展具有重要的意义。

#### 3.1 材料和方法

##### 3.1.1 研究区介绍

我国气候多样，包含了不同的气候类型，不同地区实际气候条件差异很大。而要想准确预测作物的产量，需要考虑到各个地区的不同气候条件，以及其对农作物生长的影响。例如中国南方的年降水量（大于 1200 mm）高于中国北方（小于 600 毫米），远高于中国西北部（小于 200 mm）。这意味着中国不同地区的作物可能会受到截然不同的气候条件的影响（Tao et al. 2008）。因此，为了更好地预测本研究中的作物产量，基于气候条件以及地理位置、温度、降水、植被和土壤信息将全国划分为七个子区域（赵松乔 1983）。该区域与上一章所用的分区域一致，但是由于考虑了不同的作物，区号有所改变，具体见图 3-1。其中分区 VI 不是农作物主产区，且缺乏相应的农作物产量数据，因此没有考虑该区域。更多的分区信息具体信息见（Li et al. 2019a）。图中显示了六个区域中冬小麦、春玉米、夏玉米、早稻、中稻、晚稻的站点的分布。总的来说，考虑不同的气候情况，可以帮助农民和决策者更好地了解其所在地区的气候特点，为其制定更加合理的种植策略提供依据。

### 3.1.2 数据介绍

#### 3.1.2.1 田间实验数据

作物产量数据（2013-2019 年小麦、玉米、水稻）来自国家农情监测站。取样方法：根据地块的自然分布，选取有代表性的上、中、下三类底宽选取多片，每片随机取 3-5 个地块（水稻和小麦大约  $1\text{m}^2$  [1m 宽, 1m 长], 玉米大约  $54\text{m}^2$  [3.6 m 宽, 15 m 长]），地块随机选取 3-5 个样点。作物的管理措施与当地农户水平一致。由于不是每个站点都有七年的观测数据所以冬小麦共 873 条数据，春玉米有 597 条数据，夏玉米有 509 条数据，早稻有 304 数据，中稻有 571 条数据，晚稻有 321 条数据。这些试验数据在过去 7 年没有精力重大的技术转变，因此，没有实施去趋势等方法来排除其中的人为因素的影响。

在本研究中，考虑对每种作物使用六个主要生长阶段（表 3-2）。生育期数据采集自中国气象数据共享网（<http://data.cma.cn/>）。生育期用的对应不同站点生长期的多年平均。六个生长阶段的播后天数因作物和研究区域的不同而不同（图 3-2）。一般情况下，冬小麦生长期约为 190-250 天。冬小麦播种时间为九月底至十一月初，六月初左右收获。玉米生长期约为 100-130 天。春玉米和夏玉米分别在 4-5 月和 6 月初左右播种，分别在 9 月初和 9 月下旬左右收获。早稻于 3 月下旬播种，7 月收割，生长期约 100-120 天。早稻和晚稻通常连作。因此，晚稻的种植日期是在早稻收获之后，构成约 110-125 天的生长期。中稻播种期在 4-5 月左右，生长期约 130-170 天。

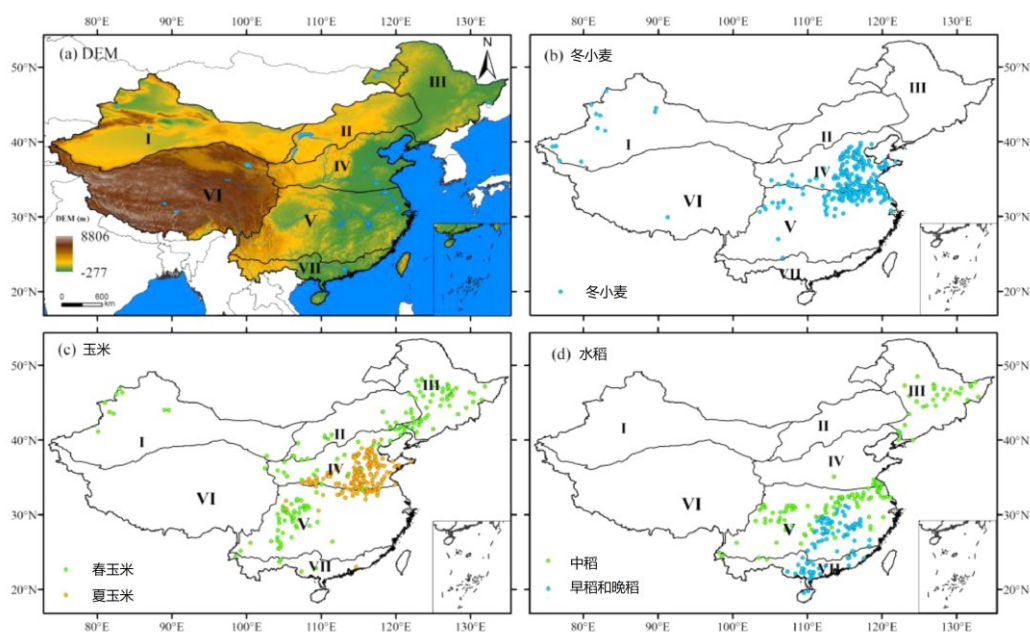


图 3-1 中国气候分区的划分(a)和中国各地冬小麦 (b)、玉米 (c) 和水稻 (d) 实验站点的分布

Fig. 3-1 Locations of seven sub-regions of China (a) and the locations of field trial sites for winter wheat (b), maize (c), and rice (d) across China

表 3-1 中全国 6 个不同子区域 6 种作物的田间试验点数量

Table. 3-1 The number of field trial sites for six types of crops in six different sub-regions of China identified in Fig. 1.

作物类型	分区					
	I	II	III	IV	V	VII
冬小麦	12	\	\	131	77	\
春玉米	10	16	50	26	48	5
夏玉米	\	\	\	123	27	\
早稻	\	\	\	\	61	32
中稻	\	\	29	8	90	4
晚稻	\	\	\	\	61	32

表 3-2 小麦、玉米、水稻六个生长阶段

Table. 3-2 Six growth-stages used for wheat, maize, and rice

Time interval	Crop		
	冬小麦	玉米 (春玉米和夏玉米)	水稻(早稻、中稻和晚稻)
T1	播种-出苗	播种-出苗	播种-出苗
T2	出苗-分蘖	出苗-三叶	出苗-分蘖
T3	分蘖-拔节	三叶-六叶	分蘖-孕穗
T4	拔节-抽穗	六叶-抽雄	孕穗-抽穗
T5	抽穗-乳熟	抽雄-乳熟	抽穗-乳熟
T6	乳熟-成熟	乳熟-成熟	乳熟-成熟



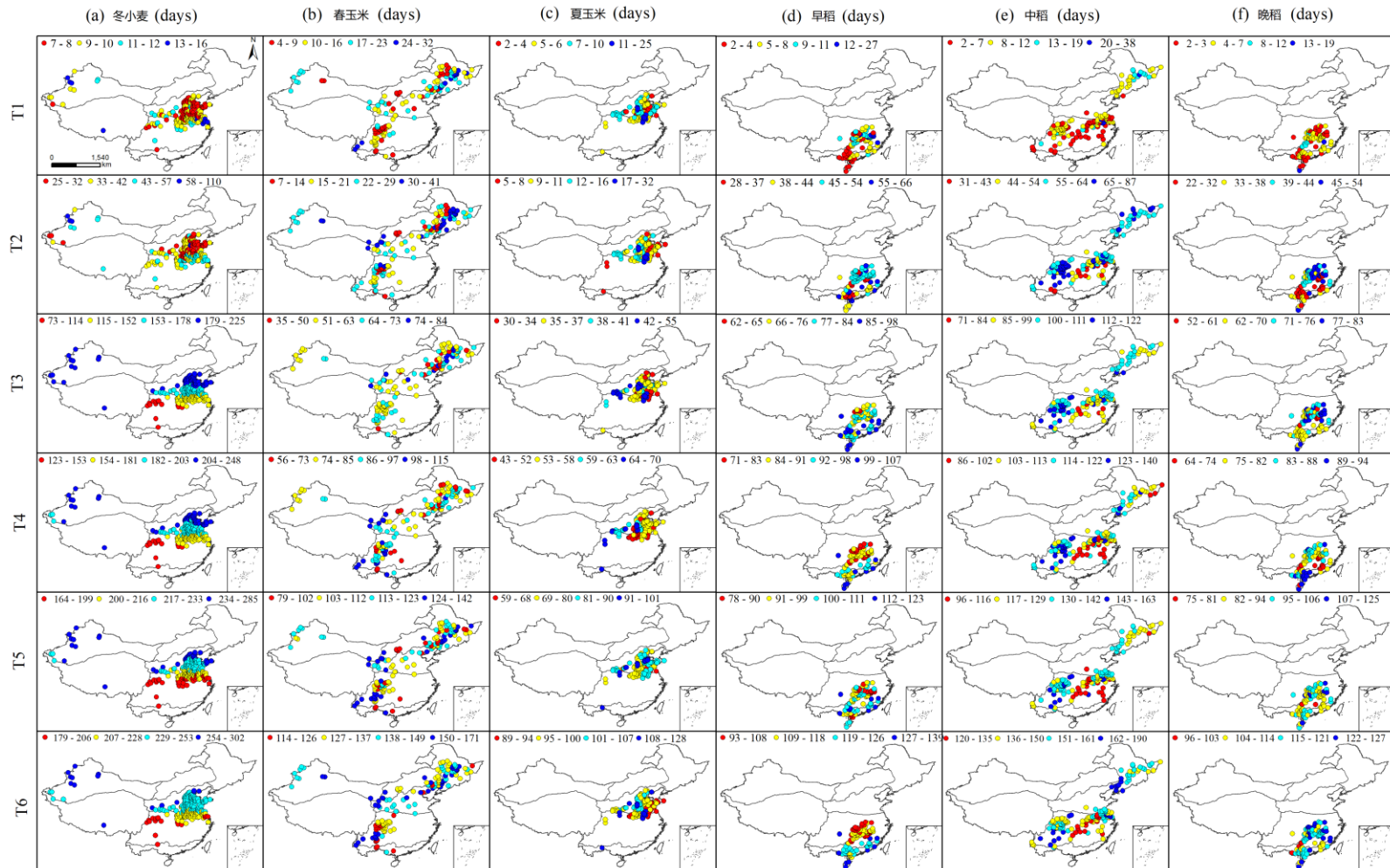


图 3-2 不同作物六个生长阶段(T1-T6)种植后天数的空间分布

Fig. 3-2 Spatial distribution of days after planting (DAP) for six growth stages (T1-T6) of different crops

### 3.1.2.2 气象数据和极端气候指数

气象数据来自中国气象数据共享网 (<http://data.cma.cn/>)。选择了离农情监测站就近的数据来反映该点的气象情况。

本研究使用了七个极端气候指数来预测作物产量，其中包括两个极端高温指数：TXx（最高温度）和 TD30（高于 30°C 的天数）；三个低温指数：TNn（最低温度）、FD0（低于 0°C 的天数，用于冬小麦）、CD10（低于 10°C 的天数，用于玉米）和 CD12（低于 12°C 的天数，用于水稻）。其中，冬小麦的霜冻日由 FD0 定义，即最低温度低于 0°C。然而，在大部分地区，玉米和水稻生长季节的最低温度都高于 0°C。因此，为了反映低温条件对玉米和水稻产量的影响，本研究定义了两个低温指数：玉米最低气温低于 12°C 的寒冷天数（CD12）和最低气温低于 10°C 的寒冷天数（CD10）。此外，还有两个极端降水指数：RX1day（最大 1 日降水量）和 SDII（降雨强度）。这些指数的来源包括 TXx、TNn、FD0、RX1day 和 SDII 来自气候变化检测和指数专家组，而 TD30 为自己定义的指数（Li et al. 2019a）。每个指标都计算了六个时段的数值，以便在每个生长阶段都能反映相应的极端气候指数（TXx、TD30、TNn、FD0（CD10 或 CD12）、RX1day 和 SDII）。

本章使用了三种不同的干旱指数：标准化降水蒸散指数（SPEI）、帕默尔干旱强度指数（PDSI）和气候水分亏缺指数（DEF）。其中，PDSI 考虑了土壤和蒸发情况，更适合揭示农业干旱状况；而 SPEI 基于水平衡，可以监测与水需求相关的干旱情况（Vicente-Serrano et al. 2010）。每月的 SPEI 被用作解释性预测因子。DEF 指数则可以有效地整合太阳辐射、蒸发蒸腾和气温对流域条件的综合影响，考虑到来自降水的可用土壤水分（Stephenson 1998）。

为了提取每个研究站点的气候数据，使用谷歌地球引擎（GEE）平台，包括月度的蒸散发量（AET）、气候水分亏缺指数（DEF）和帕默尔干旱强度指数（PDSI），并根据定义计算出每个月的站点规模 SPEI。然后，选择每个作物生长阶段最接近的月份的值作为解释性的预测因子。对于其他站点尺度的气候指数，根据指数的定义在每个站点的每个生长阶段推导出它们的值。在本章中，还使用了上一章中提到的 NDVI 和 EVI 指数，并从 GEE 上获取。

本章用了 NDVI 和 EVI 指数，从 GEE 上获取，与上一章内容一致。

### 3.1.2.3 土壤数据

土壤水分数据是从全球土地数据同化系统（GLDAS）下载的。GLDAS 使用先进的地表建模和数据同化技术，通过 GEE 平台生成地表状态和通量的最佳场 (<https://ldas.gsfc.nasa.gov/gldas/>)。本研究使用了 0-10 cm、10-40 cm、40-100 cm 和 100-200 cm 的土壤水分数据。土壤水分数据通过与测量值的比较进行评估，发现在中国各地都是准确的（Liu and Zhao 2018）。该数据集的区域分辨率为  $0.25^\circ \times 0.25^\circ$ ，



时间分辨率为 3 小时。在这项研究中，计算了每个地点每个作物生长阶段的土壤水分平均值。土壤水分数据的单位是  $\text{kg}/\text{m}^2$ 。为统一单位，将不同深度的土壤水分值换算为  $\text{m}^3/\text{m}^3(\%)$ ，计算公式如下：

$$\theta = \frac{w}{\rho \times h} \times 100\% \quad (3-1)$$

式中， $\theta$  为土壤体积含水量 (%)； $w$  为 GLDAS 土壤水分 ( $\text{kg}/\text{m}^2$ )； $\rho$  为水的密度 ( $\text{kg}/\text{m}^3$ )， $\rho = 1000 \text{ kg}/\text{m}^3$ ； $h$  为土层厚度 (m)。

### 3.1.3 模型构建

#### 3.1.3.1 建模框架

本研究基于随机森林 (RF) 构建预报模型。通过 RF 模型的“%IncMSE”评估了相对重要性。%IncMSE 表示当随机排列变量值时，在 RF 模型中使用变量的节点的准确度平均下降 (通过均方误差测试)。之前的一些研究表明，在农业研究领域，RF 模型的性能通常优于许多其他机器学习方法。

表 3-3 每个作物六个生长阶段最优的参数

Table. 3-3 The optimum values of  $m_{\text{try}}$  and  $n_{\text{tree}}$  for each RF model at six growth stages

作物	参数	T1	T2	T3	T4	T5	T6
冬小麦	$m_{\text{try}}$	19	24	30	38	48	60
	$n_{\text{tree}}$	1200	1000	600	1000	800	1000
春玉米	$m_{\text{try}}$	5	9	9	15	30	24
	$n_{\text{tree}}$	1200	1200	1000	600	1200	1200
夏玉米	$m_{\text{try}}$	7	5	9	24	15	24
	$n_{\text{tree}}$	800	1000	1200	1200	600	800
早稻	$m_{\text{try}}$	2	3	5	48	60	60
	$n_{\text{tree}}$	1000	1200	1200	1200	1200	1200
中稻	$m_{\text{try}}$	5	19	24	38	30	48
	$n_{\text{tree}}$	200	800	1200	1200	800	1200
晚稻	$m_{\text{try}}$	3	7	15	19	24	48
	$n_{\text{tree}}$	200	1000	1000	600	1200	1000

本章优化了随机森林算法的两个参数，分别是  $m_{\text{try}}$  和  $n_{\text{tree}}$ 。其中， $m_{\text{try}}$  的取值范围为 1 到变量数，步长为 1； $n_{\text{tree}}$  的取值范围为 200 到 1200，步长为 200。采用 R 语言的“tuneRF”函数来选择最优参数，该函数使用袋外 (OOB) 数据进行无偏内部验证。随机森林算法通过 bootstrap 方法构造每棵树，因此有三分之一的数据不参与树

的构造，这些数据被称为 OOB 数据。通过“tuneRF”函数，可以计算不同  $mtry$  下的 OOB 误差，选择具有相对较低 OOB 误差的参数组合，以达到更好的模型性能。具体的结果见表 3-3。

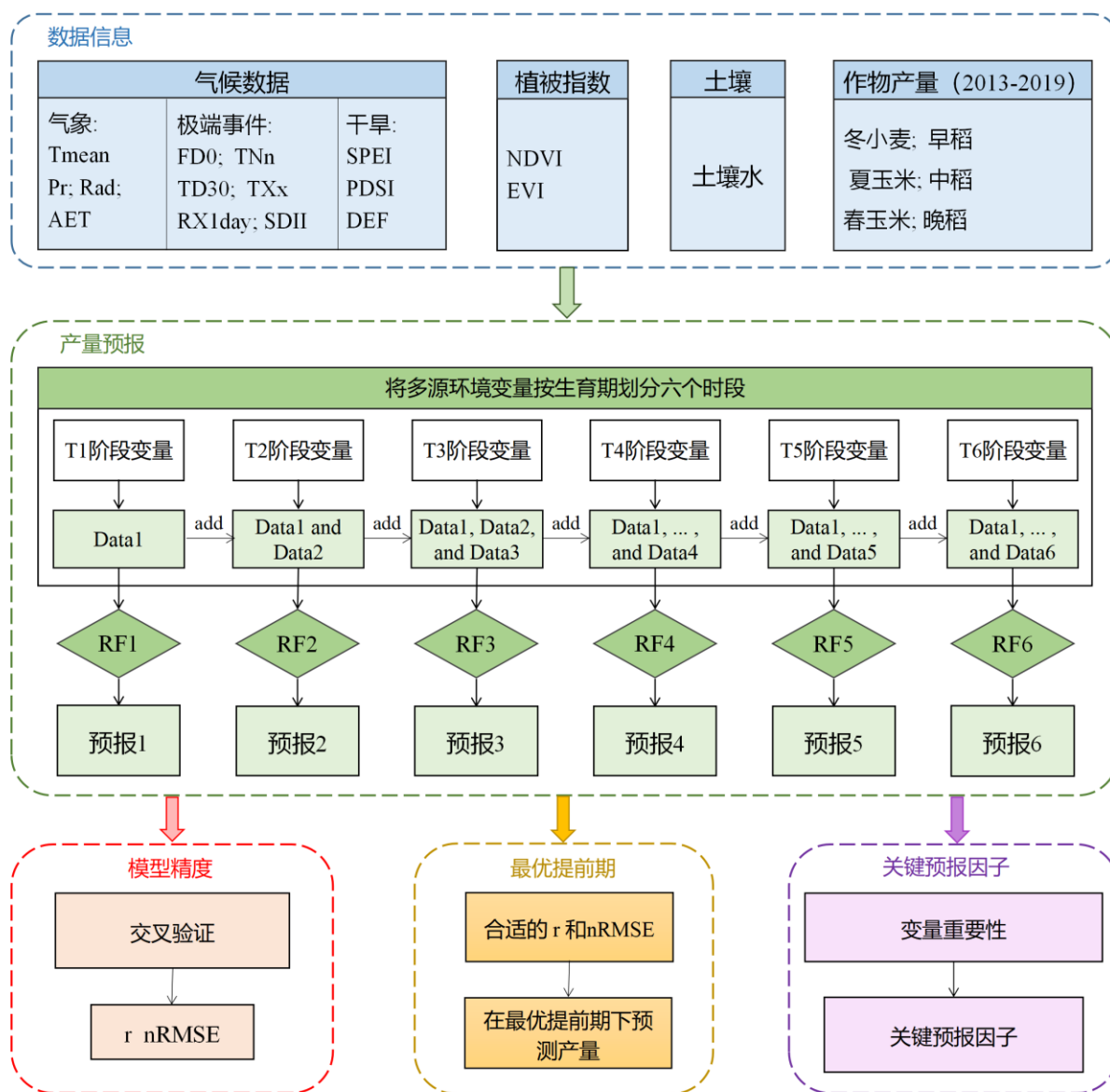


图 3-3 本章构建的不同作物产量预报模型示意图。T1-T6 代表六个生育阶段

Fig. 3-3 Schematic overview of data input and output for the RF model developed in this study. T1 to T6 represent the six growth stages for each crop type

使用 RF 算法开发了基于多源环境数据的产量预测模型。首先根据每种作物类型的不同生长阶段（从 T1 到 T6）将多源环境变量汇总为六组。然后在每个生长阶段连续触发预测事件，并在作物生长进程中添加预测因子。然后在每个生长阶段依次触发预测事件，并将每个阶段的预测因子与作物生长进程相加。因此，预测变量的数量随

着从 T1 到 T6 的进展阶段而增加。为了证明不同阶段最重要的预测因子，每个协变量的相对重要性是通过对它们在整个阶段进行平均来计算的。例如，T3 时 NDVI 的相对重要性由 T1 时的 NDVI (NDVI\_1)、T2 时的 NDVI (NDVI\_2) 和 T3 时的 NDVI (NDVI\_3) 进行平均。在本研究中，考虑了特定阶段的变量（例如，极端气候指数、植被指数和土壤水分），因此，产量预测模型可以在几个目标生长阶段（T1-T6）结束时随着生长季节的进展动态预测作物产量。在产量预测模型中，为每种作物类型开发了一个单独的 RF 模型。此外，基于 T1-T6（表 3-2）的预测阶段因作物而异（图 3-2）。整体建模框架如图 3-3 所示，展示了如何基于不同生长阶段的多个预测因子开发每个单独的 RF 模型（T1-T6，见表 3-3）。

### 3.1.3.2 模型性能评估

使用交叉验证方法评估六种作物（冬小麦、春玉米、夏玉米、早稻、中稻和晚稻）的模型性能。在本研究中，使用 Pearson 相关系数( $r$ ) 和归一化均方根误差( $nRMSE$ ) 来评估模型性能。等式写成如下：

$$r = \frac{\sum_{i=1}^n (x(i) - \bar{x})(y(i) - \bar{y})}{\sqrt{\sum_{i=1}^n (x(i) - \bar{x})^2} \sqrt{\sum_{i=1}^n (y(i) - \bar{y})^2}} \quad (3-2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x(i) - y(i))^2}{n}} \quad (3-3)$$

$$nRMSE = \frac{RMSE}{\bar{x}} \quad (3-4)$$

其中  $y(i)$  和  $x(i)$  分别是第  $i$  个预测和观察到的收益率值； $\bar{y}$  和  $\bar{x}$  代表预测值和观测值的平均值； $n$  是样本数。

在这项研究中，使用“ggplot2”R 包来制作图形。使用 ArcGIS 10.3 软件绘制田间试验点的空间分布图。

## 3.2 结果分析

### 3.2.1 模型预报精度

每个分区的小麦、玉米和水稻产量存在很大的差异，具体数据如下：冬小麦产量范围为 1952 至 8804 kg/ha，春玉米产量范围为 2997 至 17370 kg/ha，夏玉米产量范围为 4127 至 10889 kg/ha，早稻产量范围为 4607 至 8853 kg/ha，中等季稻产量为 5291

至 12185 kg/ha，晚稻产量为 4811 至 9436 kg/ha。在华北平原（IV）分区中，冬小麦单产中位数高于 I 分区和 V 分区，春玉米单产中位数从低到高分别是第 VII、第 V、第 IV、第 II、第 III、第 I 分区。在第 IV 分区，夏玉米单产中位数高于第 V 分区，而第 V 分区和第 VII 分区的早稻单产中位数相似。在第 V 分区，晚稻单产高于第 VII 分区。这种产量变化的原因可能是土壤、天气条件、干旱和极端气候事件等差异造成的。因此，基于多源环境变量预测作物产量将是一项有意义的工作。图 3-4 中的小提琴图用于说明作物产量数据的分布情况，其中的框边界表示产量的第 25 和第 75 分位数，框下方和上方的胡须表示第 10 和 90 百分位数。每个方框内的黑线表示中位数。小提琴图轮廓说明了概率密度，即阴影区域的宽度代表该产量数据的比例。

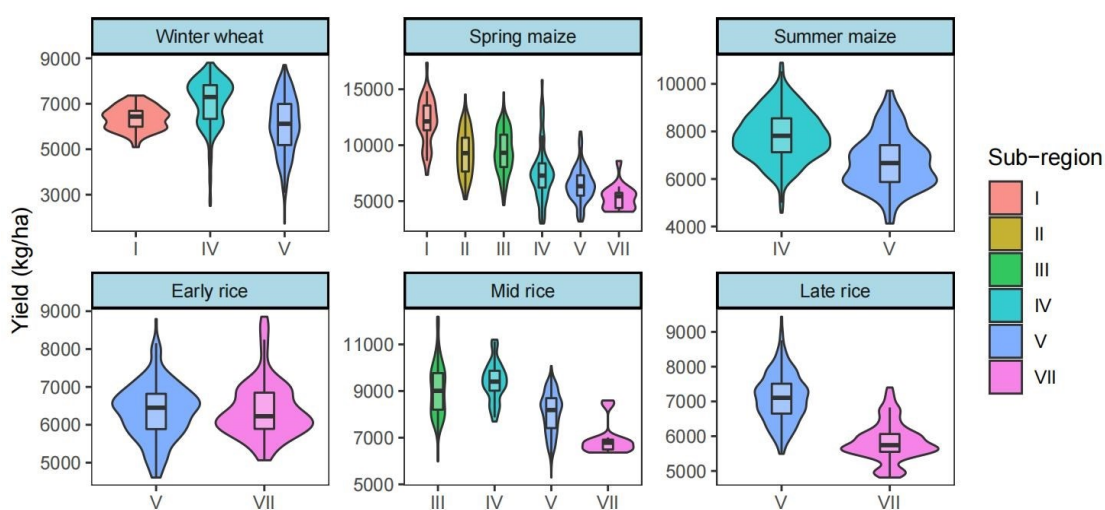


图 3-4 2013-2019 年中国六个不同子区域作物产量分布

Fig. 3-4 Distribution of measured crop yields in six different sub-regions of China during 2013-2019

为了更好地理解模型性能的稳定性，展示了六种作物类型在所有年份的不同提前期的  $r$  值和  $nRMSE$  值（见图 3-5）。模型性能通过  $r$  和  $nRMSE$  评估，图 4 显示了基于 RF 模型的六种作物预测产量的交叉验证准确性指标，包括了 2013-2019 年的所有年份。随着预测期接近生长季节结束，即  $r$  值逐渐增加， $nRMSE$  随时间减少，模型性能普遍提高。T1 时，冬小麦产量预测相对较差（ $r = 0.65$  和  $nRMSE = 14.6\%$ ），但在 T3（分蘖-拔节）及之后，预测精度大大提高。 $r$  值从 0.65（T1）增加到 0.81（T3）， $nRMSE$  从 14.6%（T1）减少到 11.4%（T3）。春玉米产量预测相对准确，因为  $r$  值从 0.71（T1）到 0.84（T6），但  $nRMSE$  值相对较高（16.2-20.4%）。夏玉米单产预测的表现从 T1（0.68）大幅上升到 T4（0.77）， $nRMSE$  从 11.3%（T1）下降到 10.4%（T4）。早稻产量预测的准确性在 T4 显着提高。预测的中稻单产  $r$  从 0.68（T1）提高到 0.84（T6）， $nRMSE$  从 9.6%（T1）降低到 7.2%（T6）。预测晚稻产量的  $r$  从

0.68 (T1) 大幅提高到 0.79 (T6), nRMSE 从 9.8% (T1) 下降到 8.4% (T6)。总的来说, RF 模型的精度比较可靠, 大多数作物在生长季节结束前的  $r$  值高于 0.75, nRMSE 低于 18.0%。因此, 结果表明, 使用多源数据的 RF 模型可以准确预测作物产量, 表明这种产量预测方法可以提供可靠的产量预测。此外, 随着每种作物的准确度从 T1 提高到 T6, 产量预测模型的稳定性也提高了。

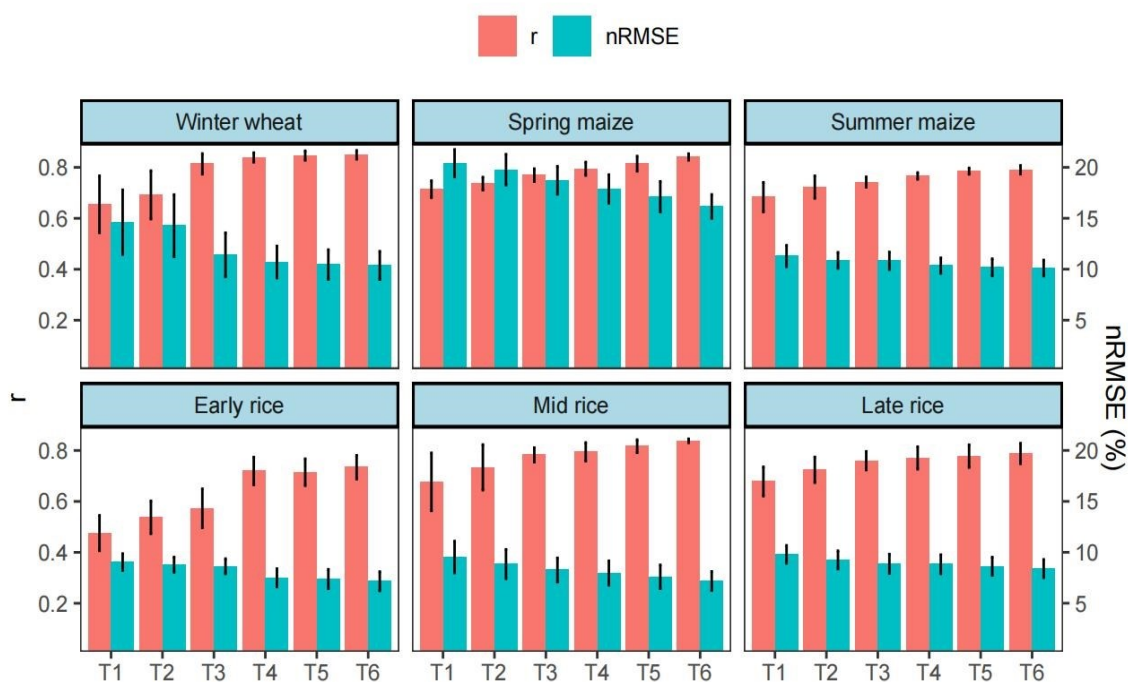


图 3-5 基于 2013-2019 年期间每年的精度平均 (逐年交叉验证), 六种作物类型在不同预测时间 (生长阶段) 的模型性能。误差棒代表七年的标准误差

Fig. 3-5 Model performance at different forecasting times (growth stages) for six crop types based on the evaluation for each (left out) year during 2013-2019. The filled bars represent the mean values of  $r$  and nRMSE; the error bars represent the standard errors for seven years

模型对每种作物不同子区域的性能如图 3-5 至图 3-7 所示。由于某些年份的采样点数量有限, 对每个子区域的所有年份 (交叉验证) 的模型性能进行了一次评估。结果表明, 即使对于同一作物, 不同子区域的模型性能也不同。这主要是由于每个子区域使用的采样点数量不同。例如, 区域 V 的夏玉米模型的  $r$  值 (在 T1-T3 期间  $r = 0.73-0.78$ ) 高于区域 IV (在 T1-T3 期间  $r = 0.58-0.65$ ), 而 nRMSE 值相似。尽管如此,  $r$  在 IV 区模型在 T4-T6 期间显著增加, 在 T6 区 IV 和 V 之间达到相似的结果。整体上, 每个模型的预测能力随着大多数子区域的生长季节向成熟而逐渐增加。

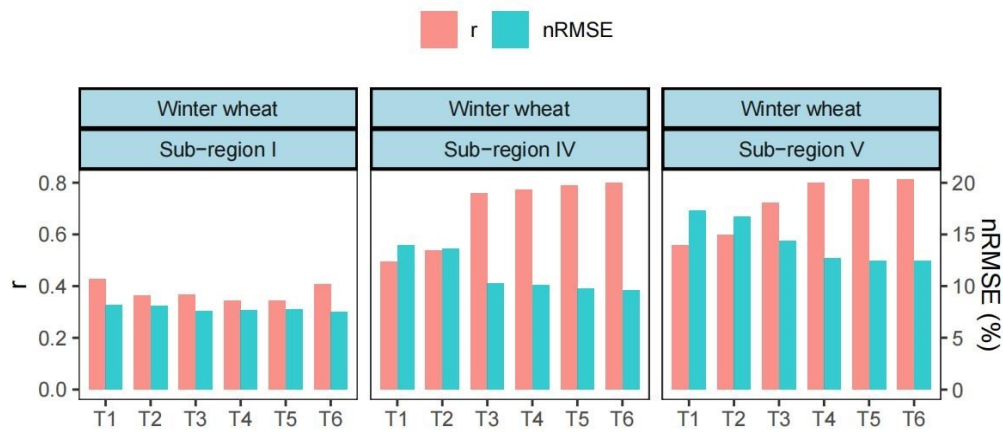


图 3-6 不同区域、生长阶段，冬小麦预报与观测的相关系数(r)与归一化均方根误差(nRMSE)

Fig. 3-6 The Pearson's correlation coefficient (r) and normalized root mean square error (nRMSE) between observed and forecasted yield for wheat in different sub-regions and growth stages

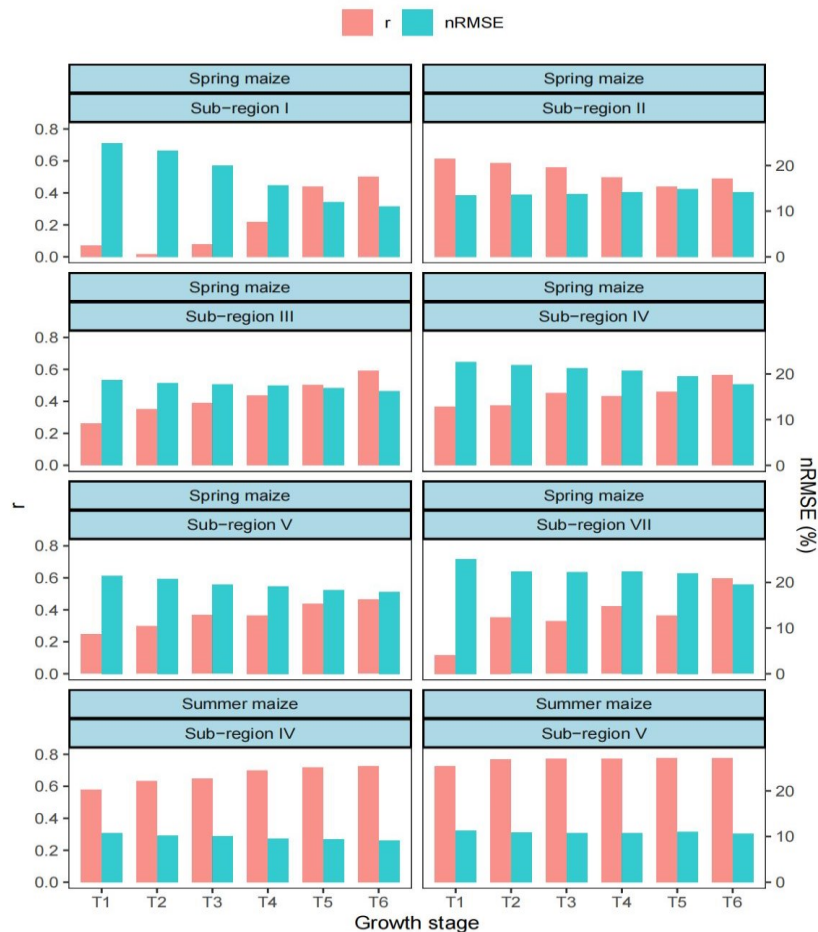


图 3-7 不同区域、生长阶段，玉米预报与观测的相关系数(r)与归一化均方根误差(nRMSE)

Fig. 3-7 The Pearson's correlation coefficient (r) and normalized root mean square error (nRMSE) between observed and forecasted yield for wheat in different sub-regions and growth stages

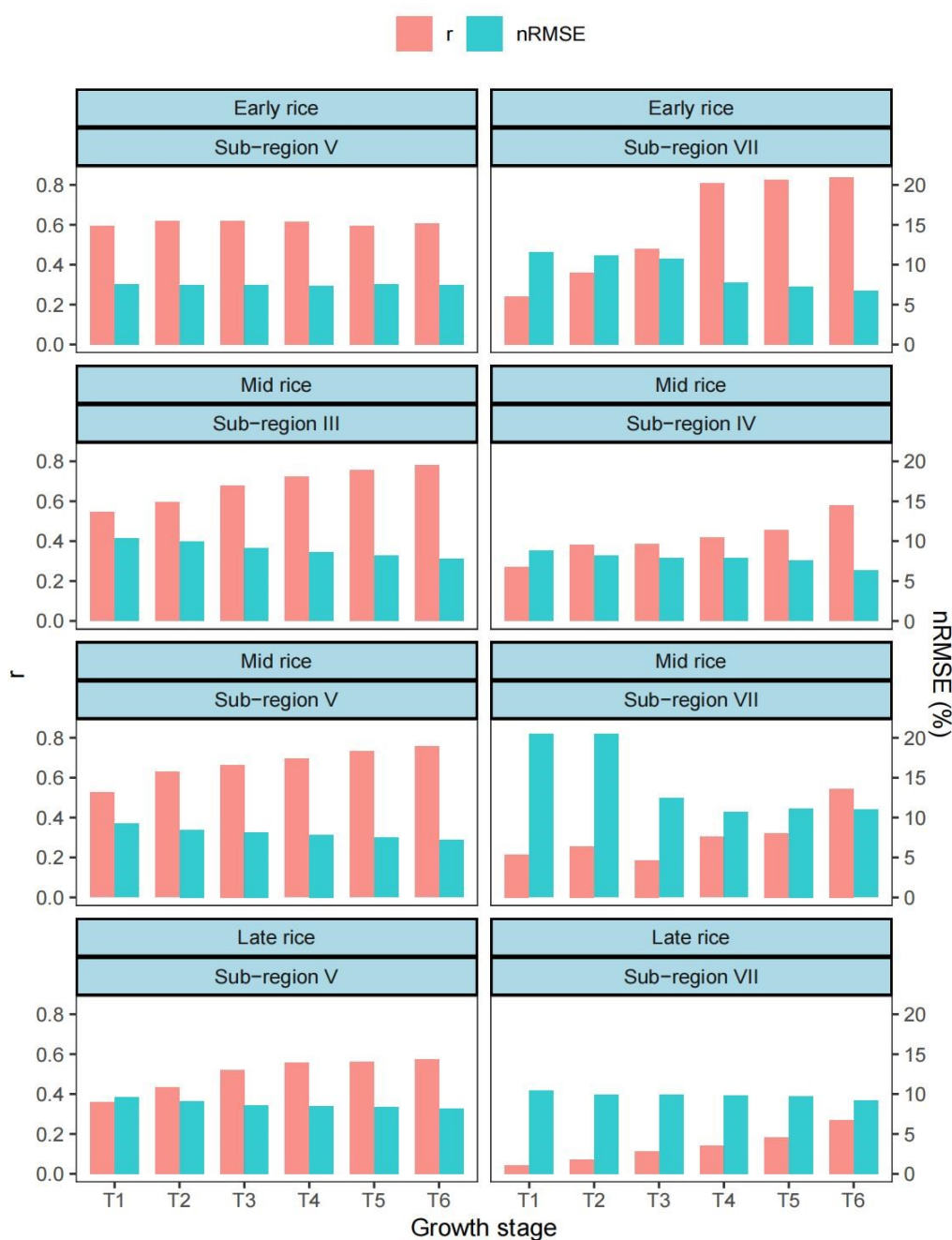


图 3-8 不同区域、生长阶段，水稻预报与观测的相关系数(r)与归一化均方根误差(nRMSE)

Fig. 3-8 The Pearson's correlation coefficient (r) and normalized root mean square error (nRMSE)

between observed and forecasted yield for wheat in different sub-regions and growth stages

### 3.2.2 产量预报的最优预报时段

根据六种作物在不同生长阶段的产量预测的准确性，可以确定产量预测系统提供的最佳提前期。总体上，预测准确率随着作物的生长发育而提高，生育后期提高速度趋缓。为了揭示在 T1 到 T6 期间观察到的每个模型性能的变化幅度，将 r 和



nRMSE 从 0% 归一化到 100% (表示为  $r_n$  和  $nRMSE_n$ ) 如图 3-8。如, 模型性能的最大增加出现在冬小麦的 T2 到 T3 期间, 其中  $r_n$  增加了 62.7%,  $nRMSE_n$  下降了 67.5% (图 5)。然而, 模型精度在 T4 之后略有提高。因此, 在最优预报时段可以选择 T4 作为一个合适的提前期。又如, 早稻预测准确率在 T4 时大幅提升 (开机抽穗,  $r_n$  提升 56.7%,  $nRMSE_n$  降低 60.5%), 而在 T4 之后, 仅略有提升 ( $r_n$  提升 5.7% 和  $nRMSE_n$  下降了 17.7%)。一般来说, 大多数作物都能达到令人满意的模型性能。

整体上, 在收获前约一到三个月, 使用 RF 模型 (图 3-9) 可以得到较为可靠的产量预报: 冬小麦 ( $r=0.81-0.85$ ,  $nRMSE=10.5-11.4%$ ); 春玉米 ( $r=0.79-0.81$ ,  $nRMSE=17.1-17.9%$ )、夏玉米 ( $r=0.77-0.79$ ,  $nRMSE=10.2-10.4%$ )、早稻 ( $r=0.71-0.72$ ,  $nRMSE=7.4-7.5%$ )、中稻 ( $r=0.78-0.82$ ,  $nRMSE=7.6-8.3%$ ) 和晚稻 ( $r=0.76-0.78$ ,  $nRMSE=8.6-8.9%$ )。

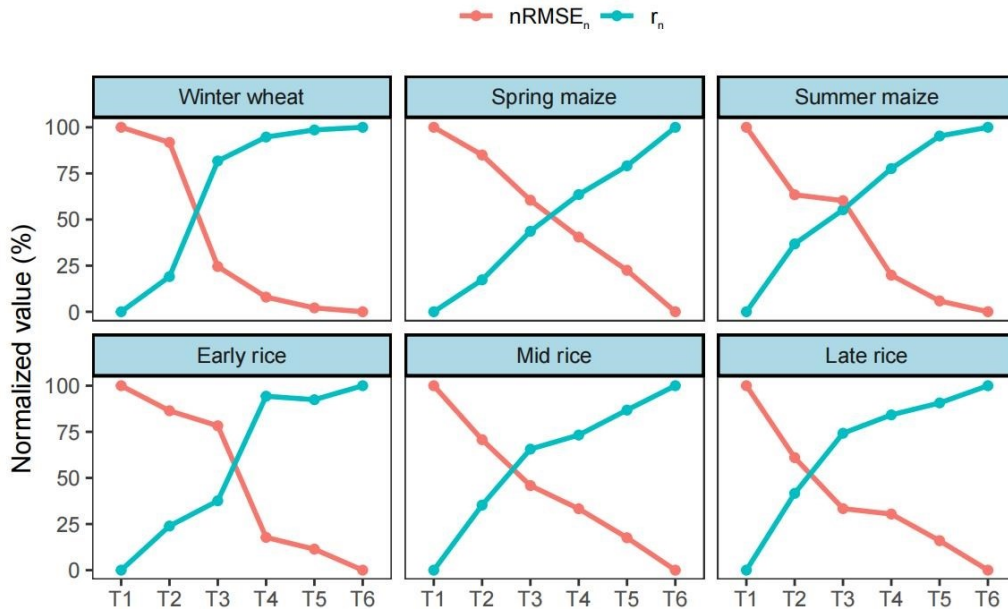


图 3-9 归一化的产量预报模型冬小麦、春玉米、夏玉米、早稻、中稻和晚稻 6 个时间区间的评价  $r_n$  和  $nRMSE_n$

Fig. 3-9 Normalized values of the model performance measures ( $r_n$  and  $nRMSE_n$ ) at six time-intervals (growth stages) for winter wheat, spring maize, summer maize, early rice, mid rice, and late rice using the RF-based forecasting model to predict crop yield in China (2013-2019).

### 3.2.2 变量重要性

每个预测变量的特征重要性用于反映不同预测变量对预测产量的贡献 (图 3-10)。一般而言, Pr、土壤和 VI 被确定为产量预测的主要预测因子。然而, 不同作物的主要预测因子各不相同。例如, 土壤水分是 T1-T2 期间冬小麦产量预测的主要预测因子。在 T3 之后, 太阳辐射 (Rad) 和 VI 是冬小麦产量预测的主要预测因子。春玉



米的主要产量预测因子是植被指数 (NDVI 和 EVI) 和干旱指数 (DI), 尤其是在 T2-T5 期间。土壤水分和 VI 是整个生长季节夏玉米产量的主要预测因子。Pr (整个生长季节, 尤其是在 T4) 和 ETH (在 T6) 是早稻产量预测的最重要预测因子。土壤水分是整个生长季节中稻和晚稻的主要预测指标。此外, 极端寒冷事件 (ETC) 也被确定为 T1-T2 期间晚稻的重要产量预测因子。在本研究中, 发现 NDVI 和 EVI 在预测水稻产量方面的重要性在水稻 (早稻、中稻和晚稻) 中排名不高。



图 3-10 6 个作物不同生长阶段的变量重要性

Fig. 3-10 The relative importance of predictor variables at different time intervals for six crop types.

图中不同的变量名: AET: 实际蒸散量; EPI: 极端降水指数 (RX1day 和 SDII); ETC: 极端低温指数 (冷日 (FD0、CD10 或 CD12) 和 TNn); ETH: 极端高温指数 (TD30 和 TXx); Rad: 短波辐射; 土壤: 土壤水分; Pr: 降水; VI: 植被指数 (NDVI 和 EVI); DI: 干旱指数 (SPEI、PDSI 和 DEF); Tmean: 平均温度。

### 3.3 讨论

#### 3.3.1 模型精度

目前有许多基于作物模型 (Li et al. 2014)、统计模型 (Cao et al. 2020) 或混合模型 (结合机器学习和作物模型) 估算作物产量 (Feng et al. 2019)。Silvestro 等 (2017) 利用遥感数据 (植被和叶面积指数等) 结合作物模型 (AquaCrop) 和卡尔曼滤波方法估算了中国西北部地区的杨陵的冬小麦产量, 模拟精度 nRMSE 为 18%。Liu 等 (2017b) 使用 DSSAT 模型估算了华北平原的玉米产量 ( $r = 0.67-0.85$  和  $nRMSE=26.8-29.8\%$ )。这些值与本研究  $r$  值相似 (春玉米为 0.84, 夏玉米为 0.79), 但在本工作中发现 nRMSE 较低 (春玉米为 16.2%, 夏玉米为 10.1%)。Guo 等 (2020) 利用使用 DSSAT 模型估算了华北平原的玉米产量 ( $r = 0.67-0.85$  和  $nRMSE=26.8-29.8\%$ )。这些值与本研究的  $r$  值相似 (春玉米为 0.84, 夏玉米为 0.79), 但在本工作中发现 nRMSE 较低 (春玉米为 16.2%, 夏玉米为 10.1%)。Son 等 (2014) 将 NDVI、EVI 和水稻产量统计数据纳入二次回归模型, 以估计越南南部春冬季水稻 ( $r = 0.79-0.84$  和  $nRMSE=6.9-8.1\%$ ) 和夏秋季水稻 ( $r = 0.63-0.75$  和  $nRMSE=5.4-6.7\%$ ), 这些值与本章的结果相似 ( $r = 0.73-0.84$  和  $nRMSE=7.2-8.4\%$ )。总之, 本研究能够在不同作物类型的产量预测方面取得与大多数先前研究相似甚至更好的性能。

此外, 结果表明模型性能因不同的子区域和作物类型而异。这一发现主要是因为本研究的产量预测系统是数据驱动的, 因此作物类型数据、数据量和数据质量是决定模型准确性的关键因素 (Peng et al. 2020a)。在本章的研究中, 试验数据集的数量、作物产量的变异性、气候条件和研究区域的大小在每个子区域都有较大差别。这些差异会影响模型在不同区域的性能。此外, 模型性能也随年份变化。T1-T3 的早稻模型性能 ( $r$ ) 在 2013-2019 年较低。这一结果主要是由于与其他作物类型相比, 早稻的试验数据集 (只有 304 条数据) 相对较少。

#### 3.3.2 小麦的产量预测

本研究显示, 在小麦产量预测中, 提前期为一到三个月可以获得令人满意的准确度 ( $r = 0.81-0.85$ ,  $nRMSE=10.5-11.4\%$ )。图 3-10 显示, T1-T2 (播种-分蘖) 阶段土壤水分是冬小麦产量的重要预测因子。这可能是由于干旱胁迫影响了小麦的潜在分蘖能力。此外, 干燥的土壤条件通常会影响到胚芽鞘和第一代分蘖的发育, 从而影响单株

分蘖数 (Blum et al. 1990)。预测准确性在 T3 之后大大提高, 这可能是因为小麦在分蘖 (例如施肥) 后生长稳定。此外, 在分蘖期, 小麦已经产生了分蘖数这一重要的产量性状, 对产量的影响比较重要 (Otteson et al. 2008)。因此, 在这些阶段, 农民应更加重视灌溉和氮肥管理, 并根据合适的目标植株密度选择最佳播种量来提高小麦分蘖数 (Bastos et al. 2020)。在 T3 之后, 辐射和植被指数被确定为决定小麦产量的主要预测因子。太阳辐射是 T3 之后特别是在 T4-T5 (拔节到乳熟发育阶段) 期间小麦产量的主要限制因素之一。

### 3.3.2 玉米的产量预测

结果表明, 春玉米( $r = 0.79-0.81$ ,  $nRMSE=17.1-17.9\%$ )和夏玉米 ( $r = 0.77-0.79$ ,  $nRMSE=10.2-10.4\%$ )的最佳提前预报期约为一到两个月。Meng 等 (2014) 基于遥感数据和回归方法预测东北地区玉米产量, 结果表明, 播种后的提前期为 55-60 天, 准确率 ( $nRMSE=7.3-16.9\%$ ) 与本研究的结果相似。然而, 他们的研究并没有考虑到干旱和极端气候事件等额外的环境因素, 且仅限于县级而非田间级别。研究结果表明, 干旱指数是影响春玉米产量的主要因素, 尤其是在抽穗期间, 这意味着该时间间隔内的干旱是对春玉米产量的主要影响。以往的许多研究也发现, 春玉米抽穗期是一个对干旱高度敏感的时期 (Atteya 2003; Gao et al. 2017; Guo et al. 2019)。在抽穗阶段补充灌溉可以增加吐丝后的耗水百分比, 从而为谷物生长发育分配更多的水和土壤氮, 最终降低产量损失的风险。此外, 本研究的结果还表明, 夏玉米的预测准确率在 T4 (六叶抽雄) 时大大提高, 在 T5-T6 期间仅有略微的提高。对这一结果的解释是, 在抽穗阶段 (T4) 之后, 冠层已经完全发育, 具有最大的辐射拦截能力 (Chen et al. 2020b)。因此, 最终产量的预测更加容易确定。

### 3.3.4 水稻的产量预测

水稻的产量预测精度在不同的提前期都可以达到可接受的水平 (图 3-9)。土壤水分是导致作物产量不稳定的主要因素之一 (Rossato et al. 2017; Singh et al. 2017)。在本研究中, 土壤水分被确定为整个生长季节中稻和晚稻产量的最重要预测因子。这是因为土壤水分不仅影响花粉和胚胎发育, 也会影响地上生长和地下生长的光合产物的转运, 特别是在灌浆期 (Chaves et al. 2002; Royo et al. 2006)。此外, 低温也被确定为晚稻产量预测的主要影响因素之一。这是因为低温可能会影响分蘖数 (T1-T2 期间), 导致花粉败育 (T3-T4 期间), 并延缓水稻生长 (Moldenhauer and Slaton 2001)。研究发现, 在 T4 (孕穗期) 时, 早稻的产量预测精度大大提高 (图 3-9)。这一结果可能是因为 T4 对气候条件非常敏感, 已被确定为决定水稻产量形成的关键阶段, 这在以往的研究中已有报道 (Asch et al. 1999; Chang et al. 2005; Fageria 2007)。因此, 生产者应优化养分管理措施, 尤其是在这些阶段。例如, Sui 等 (2013) 发现在早期营养

阶段施用高氮可以增加圆锥花序的数量。此外，调整不同生育期的施氮比例，可以改善与产量发展相关的库源问题，从而提高水稻产量。本研究的另一个重要发现是，在本研究中，植被指数对小麦和玉米产量的预测是最重要的预测因子之一，但对水稻产量预测的重要性并不大。这可能是由于未封闭的水稻冠层和土壤背景特征，这两者都会影响冠层光谱传感器，从而进一步影响用于估算水稻产量的这些植被指数(Liu et al. 2017d)。

在这项研究中，开发了一个预测中国三种主要作物产量的系统，但是也意识到这个模型存在一些局限性。首先，由于观测数据的不足，本研究无法考虑和说明田间管理方案，例如灌溉和施肥，这可能会影响产量预测的准确性。其次，品种选择也是决定作物产量的重要管理因素。不同的基因型将对不同产区的产量产生显著影响。然而，由于缺乏数据，没有在模型中包括不同的品种类型。第三，没有使用观察到的物候数据来准确地纳入作物生长阶段到本研究的模型中。Feng 等(2020)使用生物物理模型模拟作物的物候信息，并使用特定阶段的预测因子来开发机器学习模型。因此，未来可能需要开发一种混合方法，结合生物物理模型和机器学习技术，以提高作物产量预测的准确性。

### 3.4 小结

本章利用多源环境变量和中国不同田间试验点，使用机器学习方法开发了预测中国三种主要作物产量的模型。研究表明，这种由多源环境变量驱动的机器学习方法可以提供令人满意的作物产量预测，对每种作物的模型性能与之前研究的结果相似。本研究发现可以令人满意地预测冬小麦收获前一到三个月的产量 ( $r = 0.81-0.85$ ,  $nRMSE=10.5-11.4\%$ )，春玉米 ( $r = 0.79-0.81$ ,  $nRMSE=17.1-17.9\%$ )、夏玉米 ( $r = 0.77-0.79$ ,  $nRMSE=10.2-10.4\%$ )、早稻 ( $r = 0.71-0.72$ ,  $nRMSE=7.4-7.5\%$ )，中稻 ( $r = 0.78-0.82$ ,  $nRMSE=7.6-8.3\%$ )和晚稻 ( $r = 0.76-0.78$ ,  $nRMSE=8.6-8.9\%$ )。希望这些预测模型的不同提前期可以为农民和决策者提供有价值的信息，减少生长季节结束前产量损失的风险。此外，本研究还确定了影响作物产量的主要预测因子，如冬小麦的主要预测因子是太阳辐射和植被指数（特别是在拔节到乳汁发育阶段）；春玉米的主要预测因子是植被指数（整个生长季节）和干旱（尤其是在出苗到抽穗阶段）；夏玉米、晚稻和中稻的主要预测因子是土壤水分（整个生长季节）；早稻的主要预测因子是降水（特别是在孕穗期到抽穗期）。未来，将探索一种混合方法，使用生物物理模型和机器学习技术来提高作物产量预测的准确性。

## 第四章 基于多模型集合预测气候变化对作物产量的影响及不确定性分析

机器学习虽然可以较好的捕捉到产量的变化，但是机器学习无法反映作物生长发育过程，因此在预测未来气候变化对产量的影响时有较大的局限性。作物模型可以考虑气候变化对作物生长影响的过程，目前已被广泛的用来分析未来气候变化对产量影响。由于单一模型可能存在较大的偏差，很多研究用多模型集合来预测 (Asseng et al. 2014; Asseng et al. 2020; Falconnier et al. 2020)。然而，一些模型之间存在较大差异，从而弱化了每个模型的特性。同样，在预测气候变化对产量的影响时，多模型平均有时也会相互抵消，导致预测的未来产量变化几乎不变 (Swann et al. 2016; Wang et al. 2017b)，这样的结果对制定适应策略和粮食生产没有指导意义。因此，在选择模型时，要考虑其特性，以更全面地评估未来气候变化对产量的影响。

目前已有的结论是，作物模型的主要不确定性来源因地而异 (Wang et al. 2020b)。且上一章已经初步探索了不确定性来源，并发现在不同子区域，不确定性来源也有较大差别，这可能是因为它们受气候管理等因素的影响。那么不确定性来源在空间上的分布是怎样的呢？在分析不确定性来源时，模型的选择也会产生很大的影响。比如，如果随机选择了一个与实际作物较接近的作物模型，就可能低估该模型带来的不确定性。然而，如何比较模型选择对不确定性的影响呢？此外，为了全面评估作物模型的不确定性，需要模拟大量模型。但是，如何确定最小的有效模型集合数量，以减少计算量呢？

小麦、玉米、水稻和大豆是人类三分之二热量摄入的来源。本章将提出一个新的框架来预测气候变化对这些作物产量的影响，并从不同角度对产量预测的不确定性进行评估。本研究期望该研究将有助于增进对全球产量预测中不确定性的理解。

### 4.1 材料和方法

#### 4.1.1 全球作物模型

本研究使用的 GGCM 模拟器由 Franke 等 (2020a) 开发 (GGCM 阶段二)，由变化因子驱动。该方法结合了作物模型与统计模型的优点，即考虑了模型的生物物理过程也考虑了一些非线性相应，可以较好的再现全球产量变化，并且已经在过去的研究中被应用 (Liu et al. 2021a; Zabel et al. 2021)。GGCM 模拟器可以成功地捕捉作物产量对温度、降水、CO<sub>2</sub> 浓度和施氮量的响应。此外，GGCM 模拟器具有预测未来气候变化下作物产量的潜力 (Franke et al. 2020a; Rosenzweig et al. 2014b)。GGCM 提供了



近 300 万次模拟与回归。通过交叉验证，表明整体上有较好的精度。由此产生的模拟器非常灵活：它们捕捉到作物产量和产量响应的强烈地理差异，可以在对气候或 CO<sub>2</sub> 变化具有完全不同敏感性的模型上表现良好。中获取作物和特定网格单元的生长季节开始和结束日期，以确保这些数据与模拟器的假设相符。

对每个气候模式网格数据重采样为 0.5°与 GGCM 模拟器相匹配。本研究计算了生长季节温度，通过对来自每个全球气候模式（GCM）的每月温度数据进行加权平均，其中生长季节内每月的天数被用作权重。类似地，本研究也计算了生长季节降水，但是使用的是生长季节总计降水量，通过将生长季节内每月的降水总量添加在一起，并使用每月天数计算降水份额。从 GGCM 模拟阶段 1 和 2 使用的数据集 (Elliott et al. 2015; Franke et al. 2020b) 中获取作物和特定网格单元的生长季节开始和结束日期，以确保这些数据与模拟器的假设相符。模拟器可以在样本内和样本外测试中真实地再现基于过程模拟的输出。

本研究主要用了 GGCM 模拟器来分析全球气候变化对产量影响（具体介绍见第四章）。其中小麦和玉米包括九个作物模型（CARAIB、EPIC-TAMU、JULES、GEPIC、LPJ-GUESS、LPJmL、pDSSAT、PEPIC 和 PROMET）水稻和大豆包括八个作物模型（CARAIB、EPIC-TAMU、JULES、GEPIC、LPJmL、pDSSAT、PEPIC 和 PROMET）。分别在雨养的条件和灌溉条件下进行了模拟，根据 Monfreda 等（2008）的作物面积数据，汇总雨养和灌溉产量。本研究分别模拟了春小麦和冬小麦。由于 Monfreda 等（2008）没有区分冬小麦和春小麦。在本研究中根据 Müller 等（2021）的方法通过平均温度和生长季节的长度来区分冬小麦和春小麦的收获区域。然后将冬小麦产量和春小麦产量合并。模拟器使用 0.5°×0.5°全球网格上生长季节平均温度和总降水量的年值以及大气 CO<sub>2</sub> 浓度作为气候输入。因为没有考虑生长季节适应性，所以本研究仅使用 A0（无生长季节适应性）设置。

因为最新的模拟（比较阶段三）考虑了更多的参数以及情况等，把最新的模型考虑了进来与 GGCM 模拟器进行了对比。GGCM 第三阶段项目 (Jägermeyr et al. 2021) 涉及五个 CMIP6 的 GCM（GFDL-ESM4、IPSL-CM6A-LR、MPI-ESM12-HR、MRI-ESM2-0 和 UKESM10-LL）和 12 个 GGCM（ACEA、CROVER、CYGMA1p74、DSSAT-Pythia、EPIC-IIASA、ISAM、LandscapeDNDC、LPJmL、pDSSAT、PEPIC、PROMET 和 SIMPLACE-LINTUL5）。每日 GCM 数据在输入 GGCM 之前，根据 ISIMIP（www.ISIMIP.org）框架 (Cucchi et al. 2020; Lange 2019; Lange 2021) 进行偏差调整并缩小至 0.5°×0.5°全球网格。从全球平均温度对大气二氧化碳浓度增加的敏感性（即平衡气候敏感性）和其他因素的角度来看，选择这些 GCM 对整个 CMIP6 系综进行最佳采样。

### 4.1.2 气候数据

为了更全面的评估气候变化对产量的影响，本章考虑了更多的气候模式。本研究从 CMIP6 的官网上获取数据（表 4-1，<https://esgf-node.llnl.gov/search/cmip6/>）。使用双线性插值方法将该数据重新采样到  $0.5^\circ$  网格。用因子差异方法为 GGCM 仿真器生成数据。对于每个 GCM，将 1980–2010 年和 2069–2099 年间生长季平均温度和生长季总降水量的模拟差异应用于 AgMERR 的历史数据。全球平均大气  $\text{CO}_2$  含量来自国际应用系统分析研究所(<https://iiasa.ac.at/>)。2069–2099 年，SSP585 的  $\text{CO}_2$  浓度为 810 ppm，基线浓度为 360 ppm。对于给定的 GGCM，因为变化系数法使用了基准期的 AgMERRA 数据，每个 GCM 的基准产量是相同的。

### 4.1.3 聚类分析

#### 4.1.3.1 种植区域划分

根据当地作物生产力、气候条件和管理的特點，将全球种植区域划分为不同的子区域 (Chapagain et al. 2022)。使用 k 均值聚类分析方法，通过这些变量，即平均温度、太阳辐射、降水量、空气相对湿度和风速，确定了小麦、玉米、水稻和大豆的不同子区域。为了反映基准期（1980–2010 年）的作物产量，本研究使用了 9 个或 8 个 GGCM 的集合来确定历史时期的平均作物产量。选择平均温度、太阳辐射、降水量、空气相对湿度和风速作为 k 均值分析的主要气候因素。它们对作物发芽、出苗和生长速度和速率有很大影响 (Zegada-Lizarazu et al. 2012)。使用的气候数据来自 AgMERRA，这与模拟器的输入相同 (Franke et al. 2020a)。本研究使用纬度和经度来表示位置信息。根据 Müller 等 (2021)，施氮量也被视为一种管理选项。最后，为每种作物划分了 12 个子区域。

#### 4.1.3.2 模型预测聚类

本研究采用了层次聚类分析 (Morim et al. 2019) 来揭示考虑  $\text{CO}_2$  施肥时，不同集合配置下相对产量变化的差异。距离 (D) 用于计算系综成员之间的初始集群距离 (Becker et al. 1988; Borg and Groenen 2005)。本研究计算了所有涉及的网格，而不是空间平均，因此每个集群的成员在空间模式和产量变化幅度方面具有相似性。其中，小麦和玉米共有 288 名成员 (9 个 GGCM, 32 个 GCM)；水稻和大豆共有 256 个成员 (8 个 GGCM, 32 个 GCMs)。本研究将未来产量变化分为三个集群，以显示三种截然不同的产量变化预测。

表 4-1 32 个 CMIP6 气候模式的信息

Table 4-1 Information of the selected 32 GCM from CMIP6

ID	名称	机构	分辨率	ID	名称	机构	分辨率(°×°)
1	ACCESS-CM2	CSIRO	1.25 × 1.87	17	GFDL-ESM4	NOAA-GFDL	1.0 × 1.3
2	ACCESS-ESM-5	CSIRO	1.25 × 1.88	18	GISS-E2-1-G	NASA-GISS	1.0 × 1.3
3	BCC-CSM2-MR	BCC	1.12 × 1.12	19	INM-CM4-8	INM	1.5 × 2.0
4	CAMS-CSMI-O	CAMS	1.12 × 1.12	20	INM-CM5-0	INM	1.5 × 2.0
5	CanESM5	CCCma	2.8 1 × 2.81	21	KACE-1-0-G	NIMS-KMA	1.87 × 1.25
6	CAS-ESM2	CAS	1.4 × 1.4	22	MCM-UA-1-0	UA	3.75 × 2.23
7	CESM2-WACCM	NCAR	0.95 × 1.25	23	MIROC-ES2L	MIROC	2.7 × 2.8
8	CM6A-LR	IPSL	2.5 × 1.3	24	MIROC6	MIROC	1.4 × 1.4
9	CMCC-CMI-SR5	CMCC	0.9 × 1.3	25	MPI-ESM1-2-LR	MPI-M	1.87 × 1.86
10	CMCC-ESM2	CMCC	0.9 × 1.3	26	MPI-ESM1-2-HR	MPI-M	0.94 × 0.94
11	CNRM-CM6-1	CNRM-CERFACS	1.4 × 1.4	27	MRI-ESM2-0	MRI	1.125 × 1.125
12	CNRM-CM6-1-HR	CNRM-CERFACS	1.4 × 1.4	28	NESM3	NUIST	1.87 × 1.87
13	CNRM-ESM2-1	CNRM-CERFACS	1.4 × 1.4	29	NorESM2-LM	NCC	2.5 × 1.9
14	FGOALS-f3-L	CAS	1 × 1	30	NorESM2-MM	NCC	2.5 × 1.9
15	FGOALS-g3	CAS	2 × 2	31	TaiESM1	AS-RCEC	0.9 × 1.25
16	FIO-ESM-2-0	FIO-QLNM	1.25 × 0.9	32	UKESM1-0-LL	UKESM	1.3 × 1.9



#### 4.1.4 不确定性分析

##### 4.1.4.1 方差分析 (ANOVA) 方法

本研究使用方差分析方法来分析产量预测的不确定性来源。对小麦、玉米、水稻和大豆作物产量变化的不确定性进行了量化分析。方差分析的总平方和(SST)方程为:

$$SST=SS_{GGCM} + SS_{GCM} + SS_{GGCM*GCM} \quad (4-1)$$

$$P_{GGCM} (\%) = SS_{GGCM}/SST \times 100 \quad (4-2)$$

$$P_{GCM} (\%) = SS_{GCM}/SST \times 100 \quad (4-3)$$

$$P_{GGCM*GCM} (\%) = SS_{GGCM*GCM}/SST \times 100 \quad (4-4)$$

其中  $n$  为不同簇数,  $P_{GGCM}$  为 GGCM 的不确定性贡献,  $P_{GCM}$  为 GCM 的不确定性贡献,  $P_{GGCM*GCM}$  为 GCM 与 GGCM 相互作用产生的不确定性贡献。

##### 4.1.4.2 抽样方法

本研究使用方差分析来估计模型的不同样本量如何影响产量变化的总不确定性。对于作物模型, 模型选取的方式是通过  $C_i^9$  (小麦和玉米) 和  $C_i^8$  (水稻和大豆) 来选取的  $i$  是从 2-9。

$$\overline{SST}_{(i)} = \frac{1}{n} \sum_{j=1}^n (SS_{GGCM(i)} + SS_{GCM} + SS_{GGCM(i) \times GCM}) \quad (4-5)$$

$$P_{GGCM(i)} (\%) = \frac{\frac{1}{n} \sum_{j=1}^n (SS_{GGCM(i)})}{\overline{SST}_{(i)}} \times 100 \quad (4-6)$$

其中  $i$  为使用的 GGCM 编号。  $n$  为  $C_i^9$  (小麦和玉米) 和  $C_i^8$  (水稻和大豆) 的总数,  $j$  是 1, 2, ...,  $n$ 。

GCM 数量的 ANOVA 分析与 GGCM 数量的分析相似。然而, 因为 GCM 的数量是 32, 所以当 GCM 数量在 6 到 26 时, GCM 的组合将超过数千万。因此, 在对 GCM 进行采样时, 不可能使用所有组合。因此, 测试了不同大小随机子集下的不确定性变化, 发现当每种作物的大小约为 100000 ( $i$  从 6 到 27) 时, 不确定性变化趋于稳定。因此, 本研究将子采样 GCM 的最大大小设置为 100000 (图 4-1)。

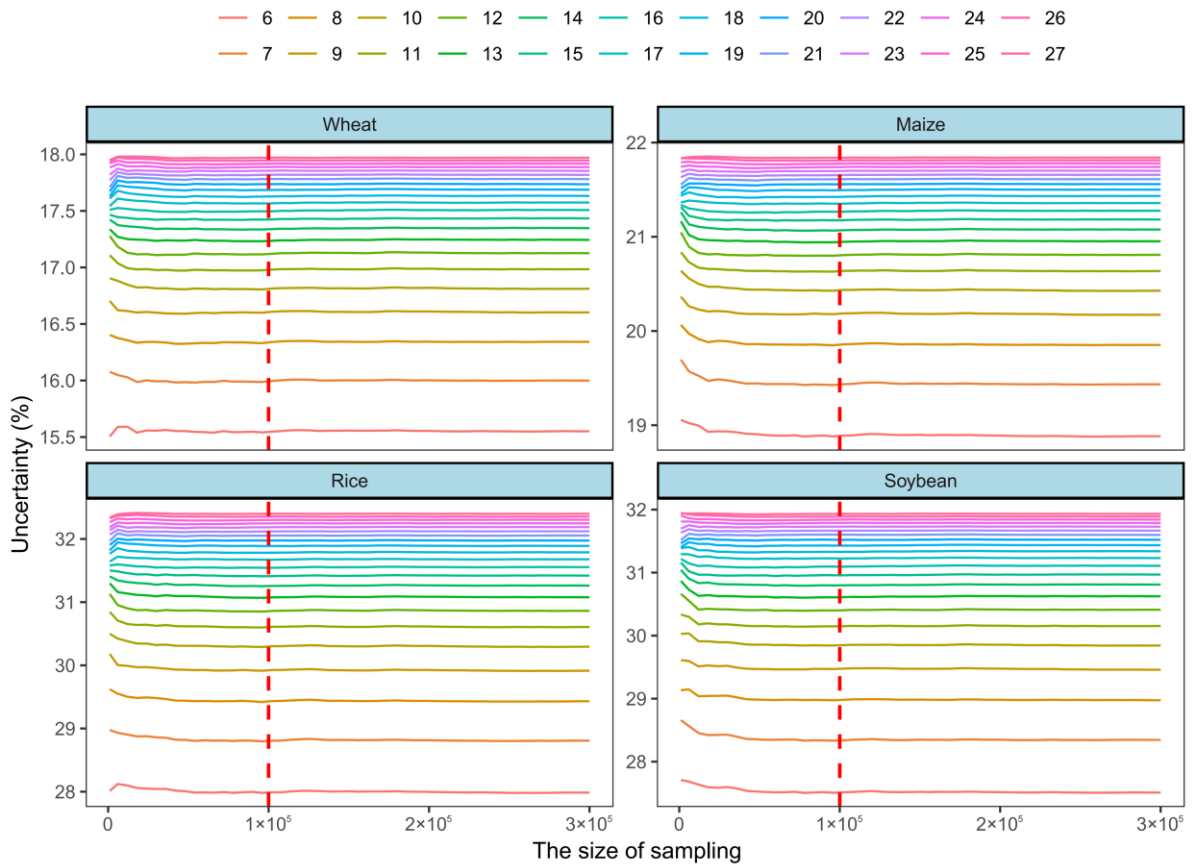


图 4-1 GCM 分别对小麦、玉米、水稻和大豆的随机抽样大小产生的不确定性。为了反映在不同数量的 GCM（从 6 到 27）下不确定性变化，本研究把随机抽样的规模设置为从 1000 到 30 万。最终使用了 100000 作为 6 至 27 的子采样 GCM 的最大大小（红色虚线），因为当大小接近该数字时，GCM 的不确定性变得稳定。

Fig. 4-1 The GCM-induced uncertainty against the size of randomly sampling for wheat, maize, rice, and soybean respectively. The size of randomly sampling under different number of GCMs (from 6 to 27) is from 1000 to 300,000 for the purpose of demonstration. We finally used 100,000 as the maximum size of subsampling GCM under 6 to 27 given that uncertainty due to GCM becomes stable when the size approaching to that number.

具体计算公式如下：

$$\overline{SST}_{(i)} = \begin{cases} \frac{1}{n} \sum_{j=1}^n (SS_{GGCM} + SS_{GCM(i)} + SS_{GGCM \times GCM(i)}) & 2 \leq i \leq 5 \text{ or } i \geq 28 \\ \frac{1}{100000} \sum_{j=1}^{100000} (SS_{GGCM} + SS_{GCM(i)} + SS_{GGCM \times GCM(i)}) & 6 \leq i \leq 27 \end{cases} \quad (4-7)$$

$$P_{GCM(i)}(\%) = \begin{cases} \frac{\frac{1}{n} \sum_{j=1}^n (SS_{GCM(i)})}{SS_{T(i)}} \times 100 & 2 \leq i \leq 5 \text{ or } i \geq 28 \\ \frac{1}{100000} \sum_{j=1}^{100000} (SS_{GCM(i)}) \times 100 & 6 \leq i \leq 27 \end{cases} \quad (4-8)$$

其中  $i$  是 GCM 的数量， $n$  是  $C_i^{32}$  的总数， $j$  是从 1 到  $n$ 。

整体上，本章具体技术路线如下：

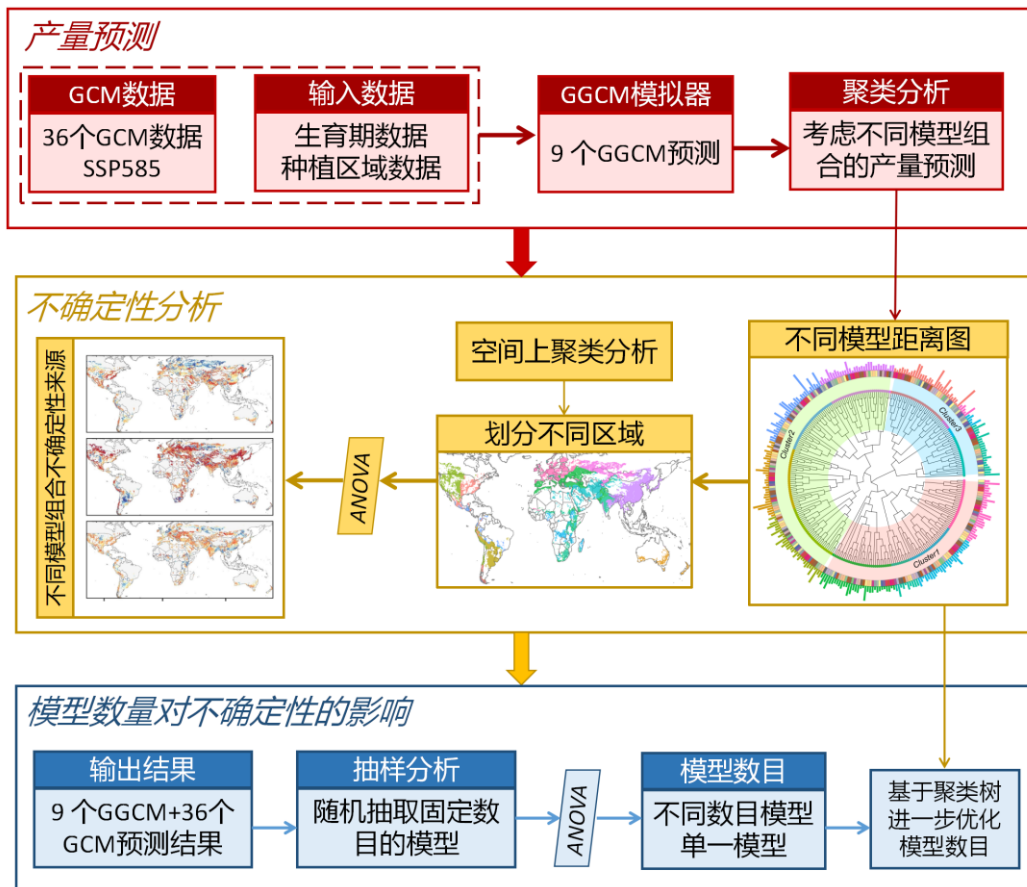


图 4-2 本章产量预测不确定性分析的技术路线图

Fig. 4-2 Schematic overview of crop yield projections uncertainty analysis in this study

## 4.2 结果分析

### 4.2.1 不同集合的产量预测

研究了 1980-2010 年和 2069-2099 年间，每个 GGCM 模拟器的小麦、玉米、水稻和大豆 GGCM 模拟作物产量的变化以及多模型集成平均值（图 4-3）。

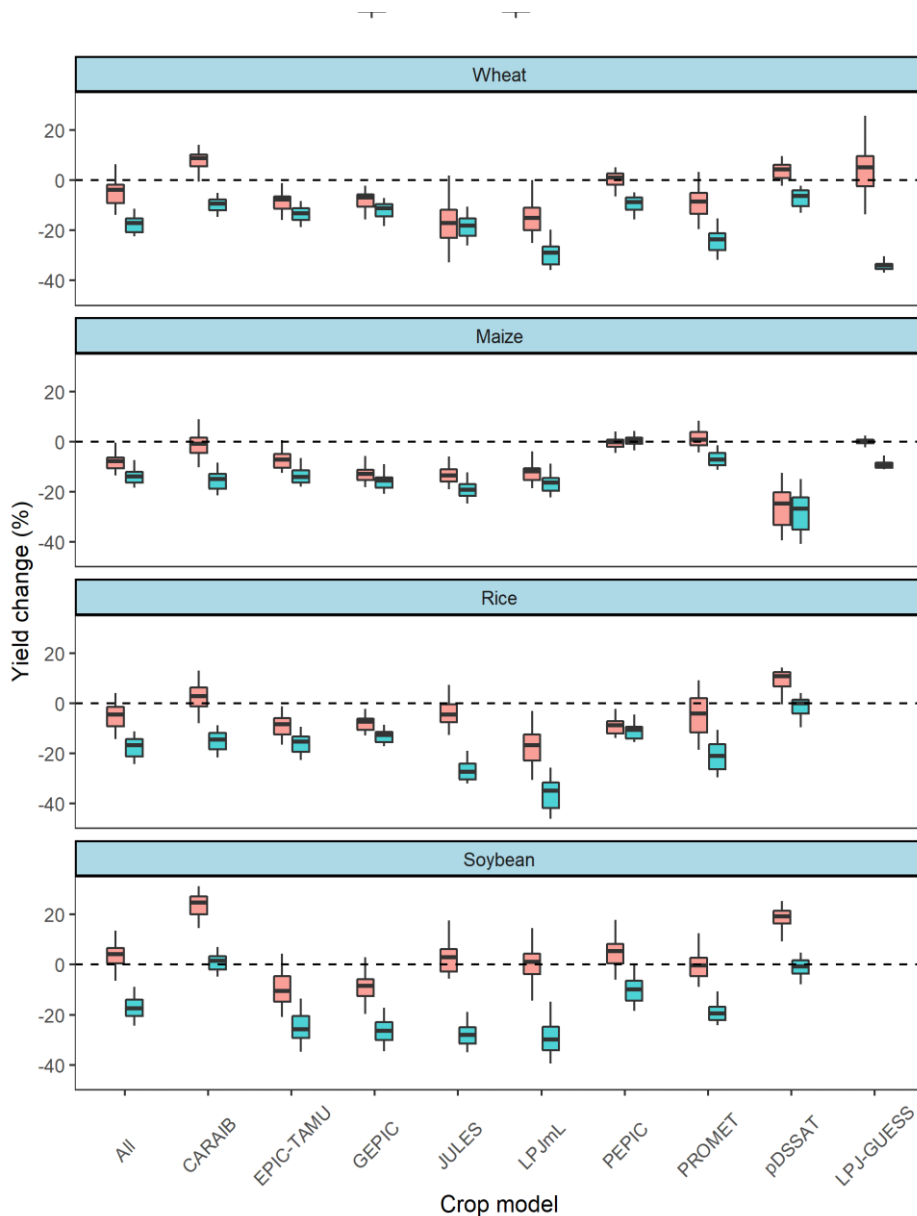


图 4-3 与基准(1980-2010 年)相比, SSP585 下多个 GGCM 和 GCM 预测的 2069-2099 期间作物产量变化(9 个 GGCM 和 32 个 GCM)。With CO<sub>2</sub>:考虑二氧化碳施肥效应; Without CO<sub>2</sub>: 没有考虑 CO<sub>2</sub> 施肥效应

Fig. 4-3 Crop yield change projected by multiple GGCMs and GCMs (9 GGCMs and 32 GCMs) under SSP585 in 2069-2099 compared to baseline (1980-2010). With CO<sub>2</sub> shows projected crop yield with the CO<sub>2</sub> fertilization effect; without CO<sub>2</sub> is projected crop yield without the CO<sub>2</sub> fertilization effect

在没有 CO<sub>2</sub> 施肥效应的情况下,大多数 GCM-GGCM 组合模拟了作物产量的下降,但有一个或两个 GGCM 模拟了玉米、水稻和大豆产量有所增加。然而,就 CO<sub>2</sub> 施肥效应而言,与小麦、水稻和大豆相比,玉米产量变化显示出更强的下降信号(见图 4-3)。此外,不同模型对 CO<sub>2</sub> 施肥效应的响应存在较大的差异,导致作物产量随 CO<sub>2</sub> 施肥

效应的变化在模型之间具有不同的表现。需要注意的是，GGCM 中实施的 CO<sub>2</sub> 施肥效应的不确定性是 GGCM 集合内方差的主要来源之一 (Müller et al. 2021)。例如，JULES 模型显示出对 CO<sub>2</sub> 施肥的较大响应，而 EPIC-TAMU 模型显示出相对较弱的 CO<sub>2</sub> 施肥效应，特别是对小麦和水稻。这些产量变化模式通常与之前使用更新的基于过程的 GGCM 的结果一致 (Jägermeyr et al. 2021)。

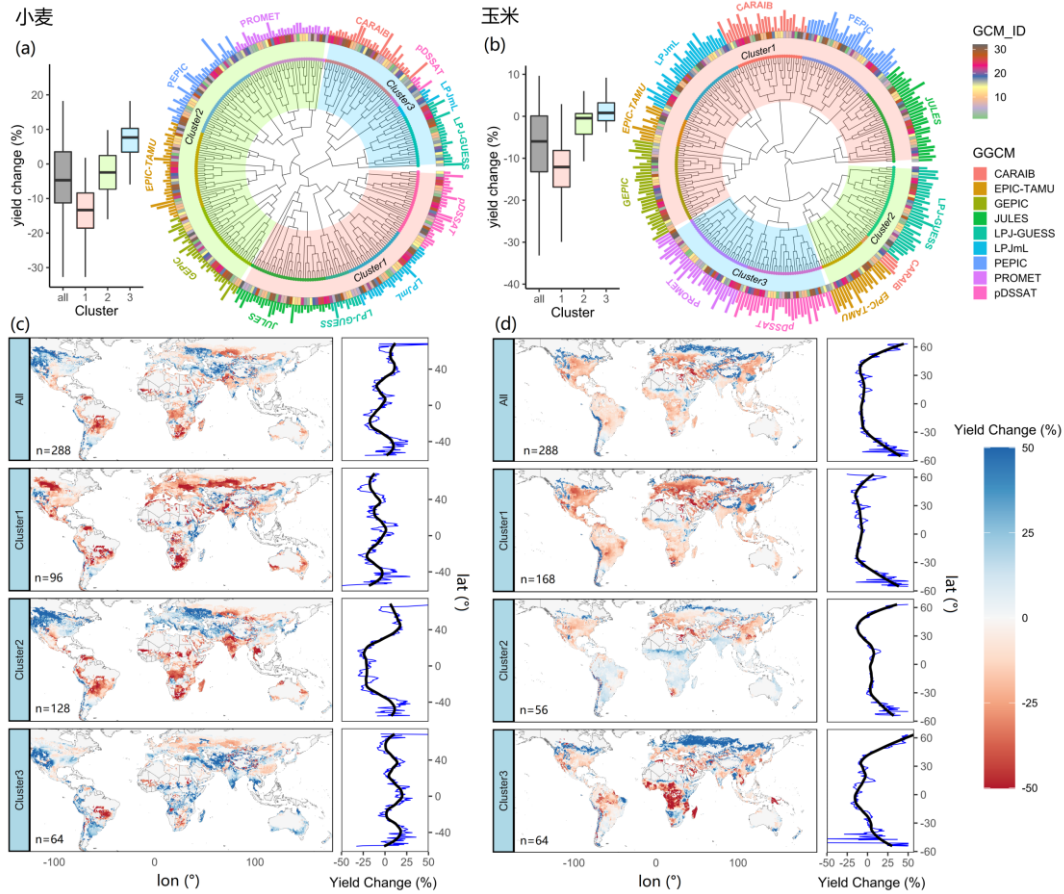


图 4-4 SSP585 下小麦和玉米产量变化(2069-2099 年和 1980-2010 年)的层次聚类 (a 和 b)。a 和 b (左) 的箱型图显示了每个集群和所有模型的全球产量变化。c 和 b (左) 是每个聚类小麦和玉米的平均产量变化。c 和 b (右) 表示产量随纬度的变化 (蓝色) 和经过平滑处理的 (黑色)。

Fig. 4-4 Hierarchical clustering of wheat and maize yield changes (between 2069–2099 and 1980–2010) under SSP585. The left panels of a and b are boxplots showing the change in global yield for each cluster and all models. The left panels of c and d are the average wheat and maize yield changes for each cluster. The right panels of c and d are the yield changes averaged by latitude (blue line) and the smoothed relationship between latitude and yield change (black line).

本研究使用聚集层次聚类分析，根据空间模式和产量变化百分比的大小，将 GCM-GGCM 组合的产量预测 (考虑 CO<sub>2</sub> 施肥) 划分为三个集群 (集群 1、集群 2 和集群 3)，并将集群 1 至集群 3 的产量变化从最低到最高排序对于四种作物 (分别为图



4-4a 和 b 以及图 4-5a 和 b)。聚类图中, 从外到内分别为: GGCM 名称、每个 GGCM 的模拟产量(条形图)、不同的 GCM、三个集群和集群树图。这三个子集分别以浅红色、绿色和蓝色显示。左边的 c 和 d 是每个集群的平均小麦和玉米产量变化。每个面板左下方的数字 (n=) 表示用于每个集群的集合成员数 (GGCM×GCM)。

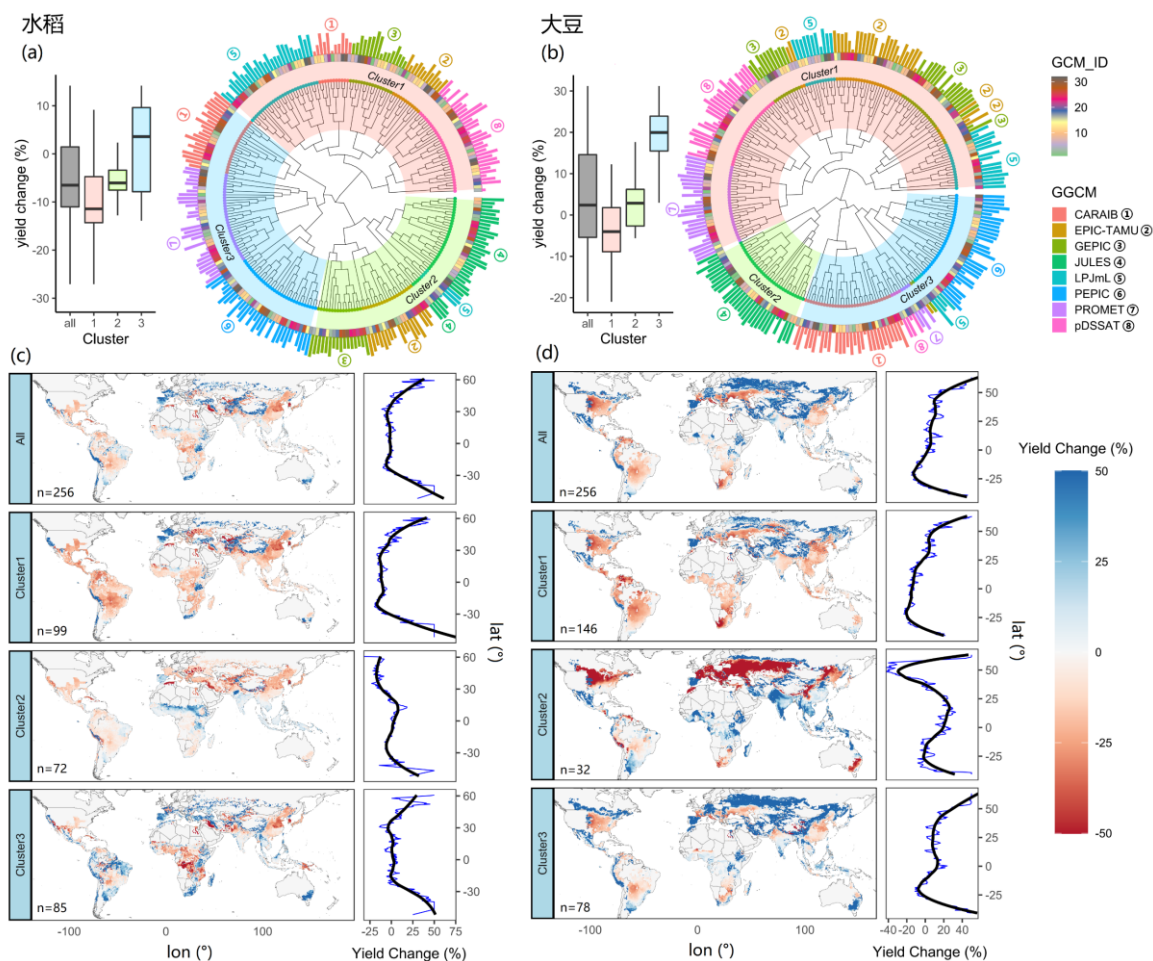


图 4-5 SSP585 下水稻和大豆产量变化(2069-2099 年和 1980-2010 年)的层次聚类 (a 和 b)。a 和 b (左) 的箱型图显示了每个集群和所有模型的全球产量变化。c 和 b (左) 是每个聚类小麦和玉米的平均产量变化。c 和 b (右) 表示产量随纬度的变化 (蓝色) 和经过平滑处理的 (黑色)。

Fig. 4-5 Hierarchical clustering of rice and soybean yield changes (between 2069–2099 and 1980–2010) under SSP585. The left panels of a and b are boxplots showing the change in global yield for each cluster and all models. The left panels of c and d are the average wheat and maize yield changes for each cluster. The right panels of c and d are the yield changes averaged by latitude (blue line) and the smoothed relationship between latitude and yield change (black line).

在一个集群内, 成员之间的产量变化幅度相似, 倾向于将它们聚集在一起形成集群。本研究将集群 1 至集群 3 的产量变化从最低到最高排序。然而, 由于不同集群之间的产量变化空间模式不同, 因此在某些特定地区, 集群 1 (3) 的产量变化并不总是

排名最低（高）（如图 4-4c 和 d 所示）。例如，在一些地区，如东非和印度南部，集群 1 呈现出积极的小麦产量响应，而集群 2 则出现了一些负面的产量变化（如图 4-4c 所示）。此外，无论是哪个集群，都会出现一些区域的产量损失，包括模拟产量有增加趋势的集群。例如，巴西南部 and 非洲南部三个集群的小麦产量均呈下降趋势（如图 4-4c 所示），表明气候变化将对这些地区产生负面影响。

由于 GGCM 第三阶段对参数以及一些设定有很大提高，为了进行对比并证明方法的可扩展性。本研究还对 GGCM 第三阶段做了类似的分析（图 4-6 至 4-9）。

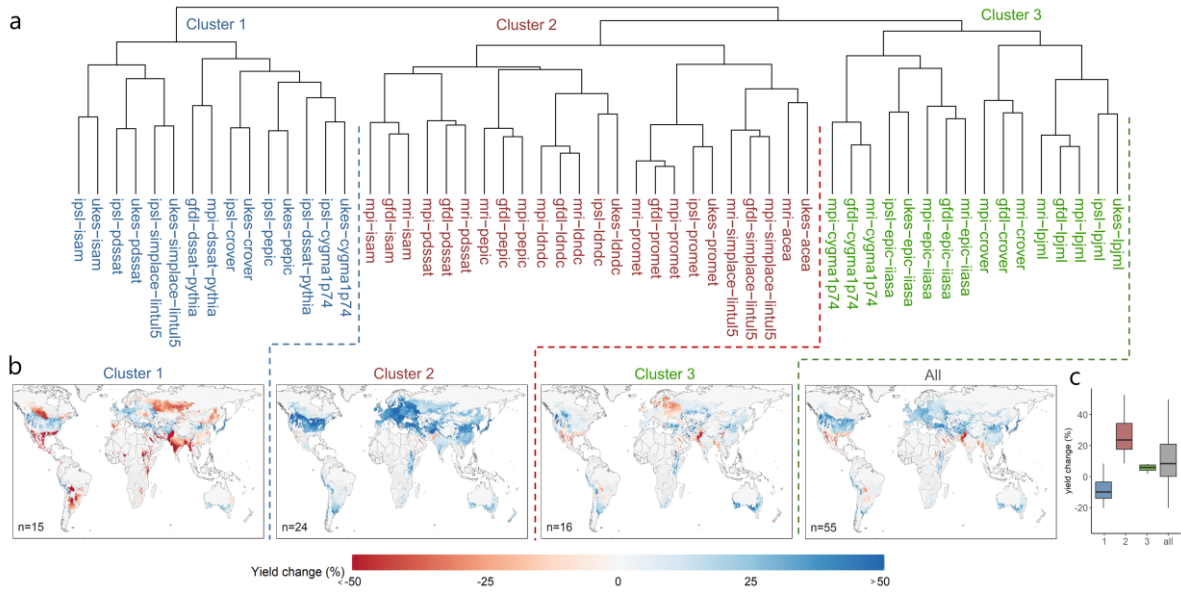


图 4-6 SSP585 下小麦产量变化的分级聚类（2069–2099 年至 1980–2010 年）。a 是小麦产量变化的聚类树图。这三个聚类分别以浅红色、绿色和蓝色高亮表示。b 是每个集群的平均小麦产量变化。每个图左下方的数字（n=）表示用于每个集群的集成员数（GGCM×GCM）。c 是显示每个集群和所有模型的全局产量变化的箱型图。边界表示作物产量变化的 25%和 75%。每个框内的黑线表示中值。gfdl、gfdl-ESM4；ipsl、ipsl-CM6A-LR；mpi、mpi-ESM1-2-HR；mri、mri-ESM2-0；ukes，UKESM1-0-LL。

Fig. 4-6 Hierarchical clustering of wheat yield changes (between 2069–2099 and 1980–2010) under SSP585. a is the cluster tree diagrams of wheat yield change. The three clusters are highlighted with light red, green, and blue. b is the average wheat yield changes for each cluster. The numbers (n=) at the bottom left of each panel represent the number of ensemble members (GGCM×GCM) used for each cluster. c is the boxplot showing the change in global yield for each cluster and all models. Box boundaries indicate the 25th and 75th percentiles of crop yield change. The black line within each box indicates the median value. gfdl, GFDL-ESM4; ipsl, IPSL-CM6A-LR; mpi, MPI-ESM1-2-HR; mri, MRI-ESM2-0; ukes, UKESM1-0-LL.

整体上，GGCM 第 3 阶段模拟生成的聚类树图显示，不同集群之间产量变化的空

间模式显示出很大的变化，这与图 4-4 和 4-5 相似。然而，GGCM 第 3 阶段的系综成员距离与 GGCM 仿真器的系综距离不同。例如，虽然 GGCM 仿真器的小麦产量变化主要由 GGCM 为核心聚类（图 4-4a），但 GGCM 阶段 3 集群 1 中的小麦产量改变主要由两个 GCM（IPSL-CM6A-LR 和 UKESM1-0-LL）分类（图 4-6）。这可以归因于这两种 GCM 表现出更高的平衡气候敏感性（ECS）（Bonnet et al. 2021; Galik 2019），这导致一些气候敏感作物模型的预期产量大幅下降。然而，值得注意的是，一些作物模型（如 DNDC、PROMET 和 LPJml）仍然属于五个 GCM 中的同一分支（图 4-6a），这表明不同的作物模型对气候变化的敏感程度不同。对于玉米来说（图 4-7），不同集合之间差异更大，例如子集 1 中产量为上升，然而子集 2 中产量为大幅度下降。

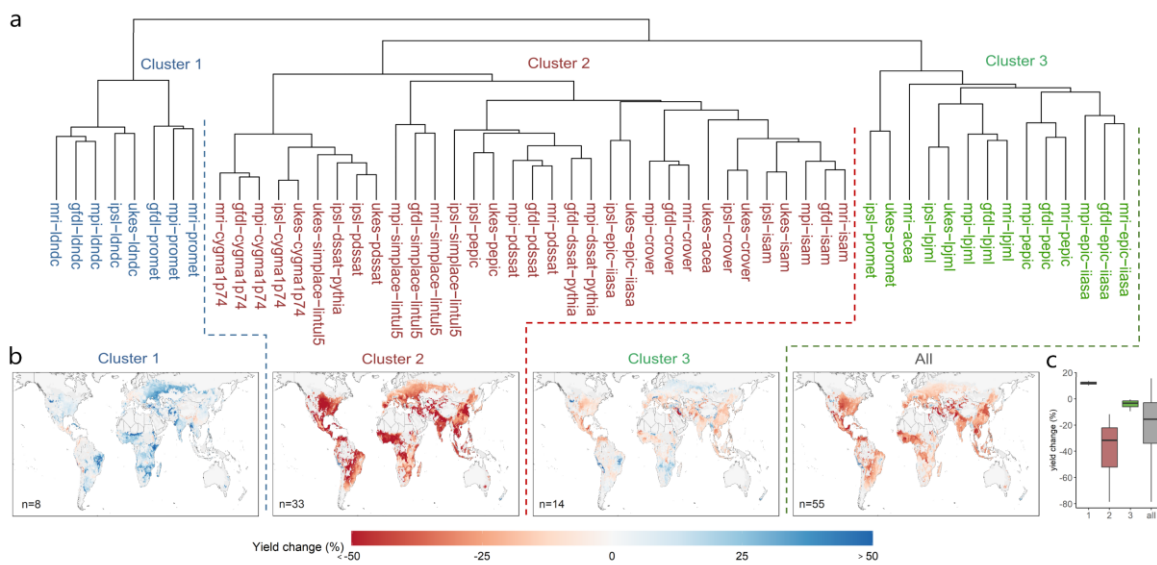


图 4-7 SSP585 下玉米产量变化的分级聚类（2069–2099 年至 1980–2010 年）。a 是小麦产量变化的聚类树图。这三个聚类分别以浅红色、绿色和蓝色高亮表示。b 是每个集群的平均小麦产量变化。每个图左下方的数字（n=）表示用于每个集群的集成员数（GGCM×GCM）。c 是显示每个集群和所有模型的全局产量变化的箱型图。边界表示作物产量变化的 25% 和 75%。每个框内的黑线表示中值。gfdl、GFDL-ESM4；ipsl、IPSL-CM6A-LR；mpi、MPI-ESM1-2-HR；mri、MRI-ESM2-0；ukes、UKESM1-0-LL。

Fig. 4-7 Hierarchical clustering of maize yield changes (between 2069–2099 and 1980–2010) under SSP585. a is the cluster tree diagrams of wheat yield change. The three clusters are highlighted with light red, green, and blue. b is the average wheat yield changes for each cluster. The numbers (n=) at the bottom left of each panel represent the number of ensemble members (GGCM×GCM) used for each cluster. c is the boxplot showing the change in global yield for each cluster and all models. Box boundaries indicate the 25th and 75th percentiles of crop yield change. The black line within each box indicates the median value. gfdl, GFDL-ESM4; ipsl, IPSL-CM6A-LR; mpi, MPI-ESM1-2-HR; mri, MRI-ESM2-0; ukes, UKESM1-0-LL.



对于水稻和大豆，也可以看到类似的规律（图 4-8 和图 4-9）。另外，由于水稻和大豆不同集合之间的差异更大，这导致了它们的多模型平均相互抵消。这也是本研究想强调的：使用多个作物模型集合可以提高产量变化预测的准确性，但同时也存在一些局限性。首先，多个作物模型集合由于存在各自的不确定性和误差，可能会抵消某个作物模型本身的过程，导致预测未来产量变化较小。其次，多个作物模型集合缺乏指导意义，即难以明确说明应采取何种措施来应对可能的产量变化，从而限制了该方法在实际应用中的可操作性。最后，多个作物模型集合可能会低估气候变化的影响，导致对未来产量变化的预测结果过于保守，从而影响决策者对应对气候变化的行动计划的制定。因此，在使用多个作物模型集合时，需要充分认识其局限性，并进行有效的解释和分析。

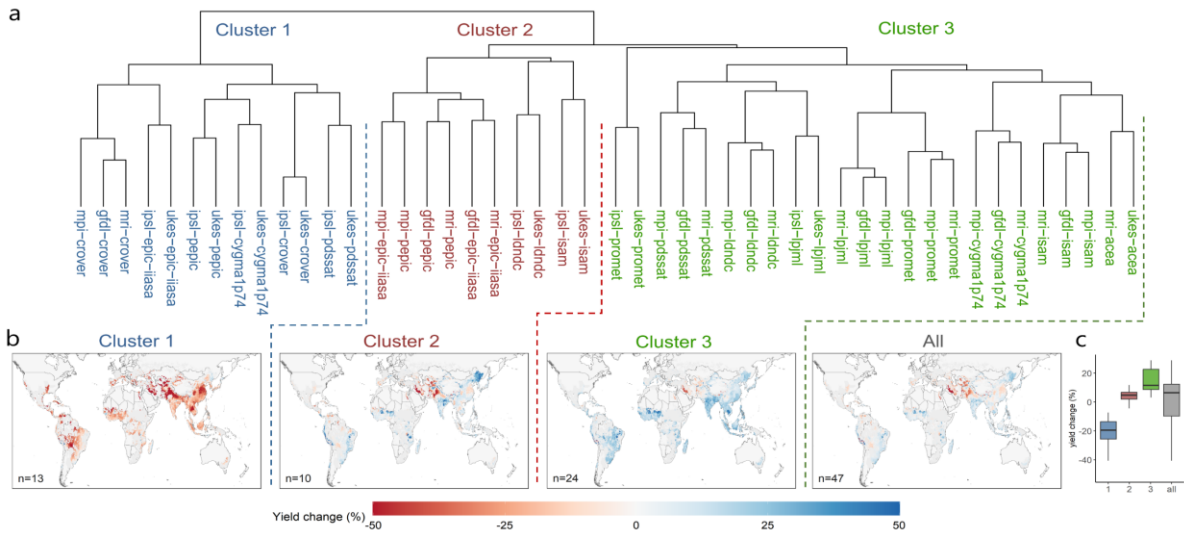


图 4-8 SSP585 下水稻产量变化的分级聚类（2069–2099 年至 1980–2010 年）。a 是小麦产量变化的聚类树图。这三个聚类分别以浅红色、绿色和蓝色高亮表示。b 是每个集群的平均小麦产量变化。每个图左下方的数字（n=）表示用于每个集群的集成员数（GGCM×GCM）。c 是显示每个集群和所有模型的全局产量变化的箱型图。边界表示作物产量变化的 25%和 75%。每个框内的黑线表示中值。gfdl、gfdl-ESM4；ipsl、ipsl-CM6A-LR；mpi、mpi-ESM1-2-HR；mri、mri-ESM2-0；ukes、UKESM1-0-LL。

Fig. 4-8 Hierarchical clustering of rice yield changes (between 2069–2099 and 1980–2010) under SSP585. a is the cluster tree diagrams of wheat yield change. The three clusters are highlighted with light red, green, and blue. b is the average wheat yield changes for each cluster. The numbers (n=) at the bottom left of each panel represent the number of ensemble members (GGCM×GCM) used for each cluster. c is the boxplot showing the change in global yield for each cluster and all models. Box boundaries indicate the 25th and 75th percentiles of crop yield change. The black line within each box indicates the median value. gfdl, GFDL-ESM4; ipsl, IPSL-CM6A-LR; mpi, MPI-ESM1-2-HR; mri, MRI-ESM2-0; ukes, UKESM1-0-LL.

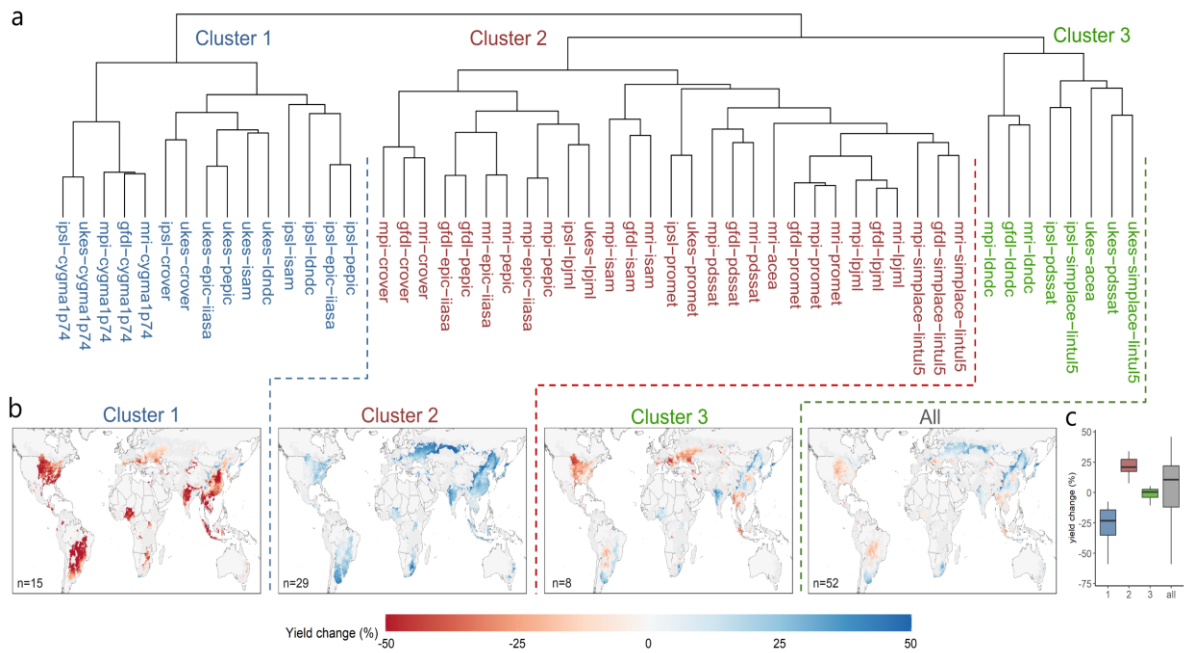


图 4-9 SSP585 下大豆产量变化的分级聚类 (2069–2099 年至 1980–2010 年)。**a** 是小麦产量变化的聚类树图。这三个聚类分别以浅红色、绿色和蓝色高亮表示。**b** 是每个集群的平均小麦产量变化。每个图左下方的数字 ( $n=$ ) 表示用于每个集群的集合成员数 (GGCM $\times$ GCM)。**c** 是显示每个集群和所有模型的全局产量变化的箱型图。边界表示作物产量变化的 25% 和 75%。每个框内的黑线表示中值。gfdl、gfdl-ESM4; ipsl、ipsl-CM6A-LR; mpi、mpi-ESM1-2-HR; mri、mri-ESM2-0; ukes, UKESM1-0-LL。

Fig. 4-9 Hierarchical clustering of soybean yield changes (between 2069–2099 and 1980–2010) under SSP585. a is the cluster tree diagrams of wheat yield change. The three clusters are highlighted with light red, green, and blue. b is the average wheat yield changes for each cluster. The numbers ( $n=$ ) at the bottom left of each panel represent the number of ensemble members (GGCM $\times$ GCM) used for each cluster. c is the boxplot showing the change in global yield for each cluster and all models. Box boundaries indicate the 25th and 75th percentiles of crop yield change. The black line within each box indicates the median value. gfdl, GFDL-ESM4; ipsl, IPSL-CM6A-LR; mpi, MPI-ESM1-2-HR; mri, MRI-ESM2-0; ukes, UKESM1-0-LL.

本研究提出了一种全新的集成数据分析方法，该方法通过定义不同集成成员之间的距离度量，并将大集成减少为三个关键模式。这种方法不仅提供了对设计集成配置更深入的理解，而且避免了多模型集成可能模糊结果并忽略重要细节的问题。此外，本研究的集群方法还有助于提高作物模型和全球气候模型之间的协调性。例如，可以通过确定敏感作物模型（例如，对气候变暖敏感的作物模型）是由高 ECS GCM 驱动的，从而预测这些模型将比其他作物模型（如图 4-6、图 4-8 和图 4-9 中的集群 1）经历更大的减产。这一方法不仅有助于预测未来作物产量的变化趋势，还可以帮助决策

者制定应对气候变化的行动计划。

### 4.2.2 不确定性分析

作物产量模拟受到当地环境和管理实践（如气候条件和作物管理）的影响，这些实践进一步影响产量变化预测的不确定性来源。因此，使用了气候变量（生长季节平均温度、降水量、风速、太阳辐射和相对湿度）、历史平均产量、氮肥施用率、经度、将全球种植区域划分为每种作物的 12 个子区域（图 4-10 为小麦和玉米；图 4-11 为水稻和大豆）。

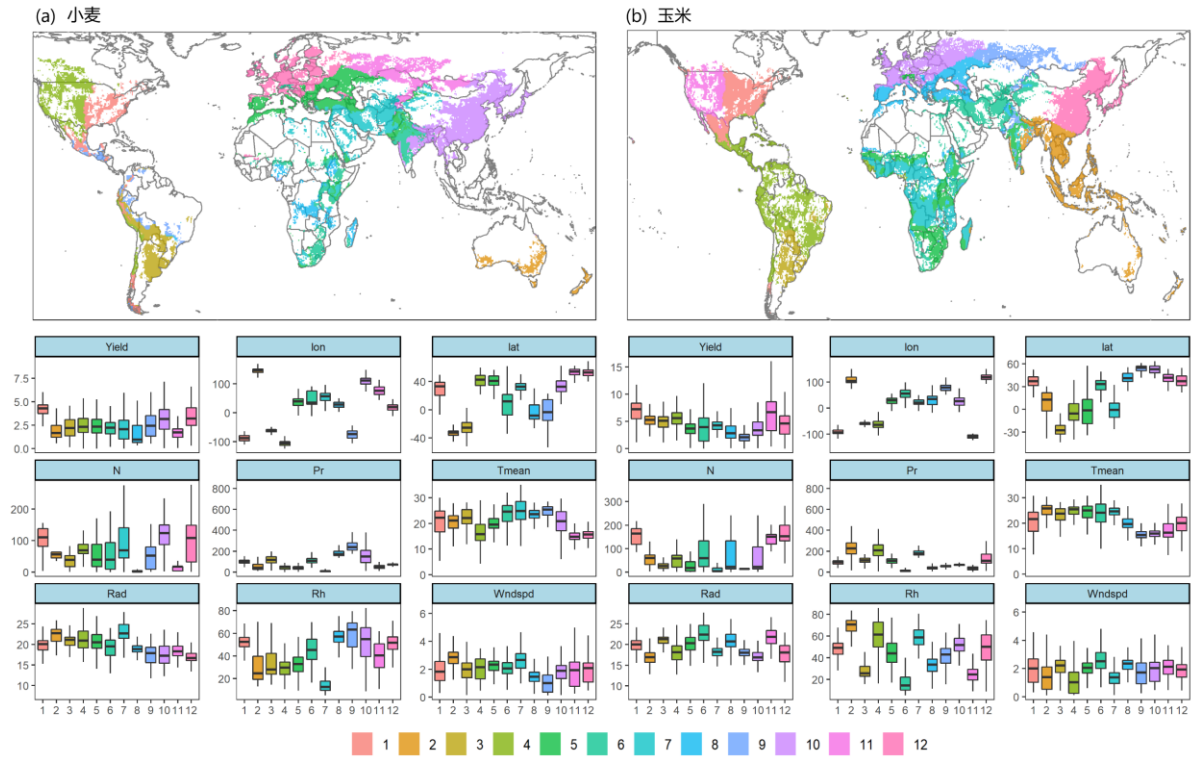


图 4-10 小麦(a)和玉米(b)基于不同变量和不同变量范围的聚类划分的空间格局（上）以及箱型图所展示的变量的范围（下）

Fig. 4-10 Spatial patterns of cluster divisions (top) based on different variables and the range of different variables (boxplots on the bottom) for wheat (a) and maize (b).

图中，Yield，作物产量(kg/ha)；经度(°)；纬度(°)；N，施氮量(kg/ha)；Pr，降水量(mm)；Tmean，温度(°C)；Rad，辐射(MJ/m<sup>2</sup>)；相对湿度(%)；Wndspd，风速(m/s)。使用 k-means 聚类生成 12 个子区域(编号为 1-12，不同颜色)。方框边界表示每个变量的第 25 和第 75 个百分点，方框下方和上方的胡须表示第 10 和第 90 个百分点。每个方框内的黑线表示中值。

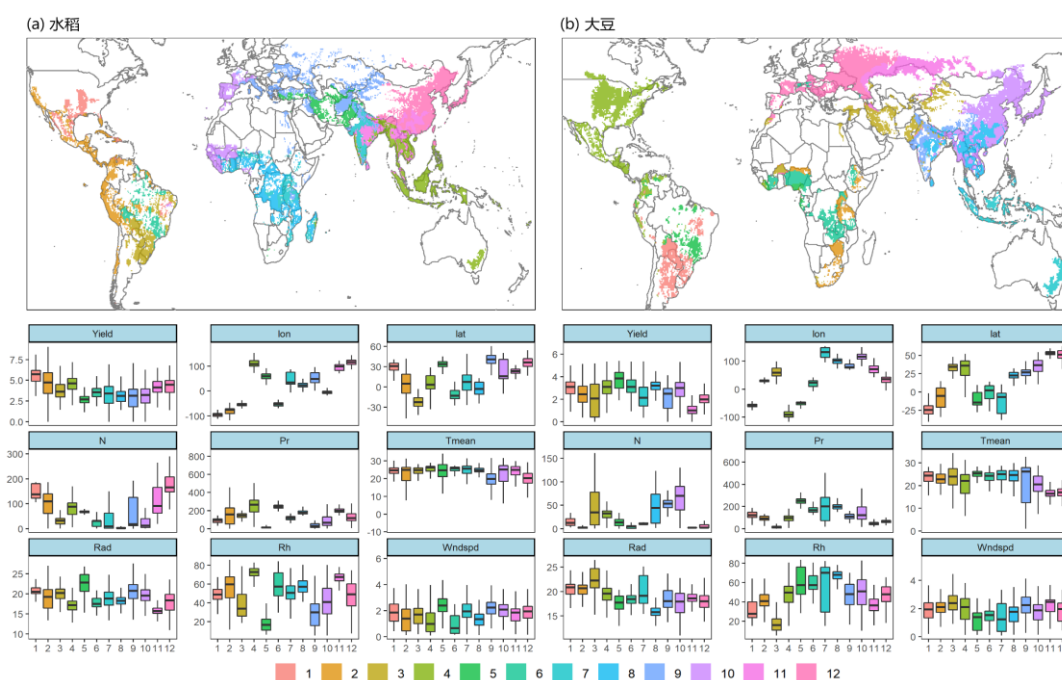


图 4-11 水稻(a)和大豆(b)基于不同变量和不同变量范围的聚类划分的空间格局（上）以及箱型图所展示的变量的范围（下）

Fig. 4-11 Spatial patterns of cluster divisions (top) based on different variables and the range of different variables (boxplots on the bottom) for rice (a) and soybean (b).

本研究对基于聚类分析的 GGCM 和 GCM 组合的每个子集在网格、纬度和区域尺度下分别分析了不确定性的主要来源，并在图 4-12 中进行了展示。整个 GCM 和 GGCM 集合中，作物模型引起的方差比 GCM 更大（图 4-12a 和 c，左图）。对于不同的集群，不同作物产量变化的主要不确定性来源存在差异，尤其是玉米、水稻和大豆。这些结果提供了不同模型组合如何影响不确定性建模来源的实例。例如，玉米的集群 2 ( $n=56$ ,  $n_{\text{GCM}}=2$ , 图 4-4d) 和集群 3 ( $n=64$ ,  $n_{\text{GCM}}=3$ , 图 4-12d) 的总体大小相似，但不确定性的主要来源在不同的空间尺度上表现出很大的差异（图 4-12c 和 d）。在集群 2 中，GCM 是北欧、东亚和美国的主要不确定性来源（图 4-12c）。而在集群 3 中，GGCM 是这些地区不确定性的主要来源。此外，水稻的聚类 2 ( $n=72$ ,  $n_{\text{GCM}}=3$ , 图 4-5) 和聚类 3 ( $n=85$ ,  $n_{\text{GCM}}=3$ , 图 4-5) 的总体大小具有相同数量的 GGCM，但它们显示了不同子区域的不确定性来源的差异（图 4-13）。这是因为，与集群 3 相比，集群 2 的三个 GGCM 具有相似的生理过程（注意 EPIC-TAMU 和 GEPIC 模型均由 EPIC 开发）。一些研究只使用了少量作物模型（约三种作物模型）分析了不确定性来源，因此，如果使用类似过程的作物模型，可能会影响结果的稳健性（Huang et al. 2022; Xiong et al. 2019）。因此，本研究的结果强调了在考虑不同模型组合时的重要性，特别是在农业影响研究中。



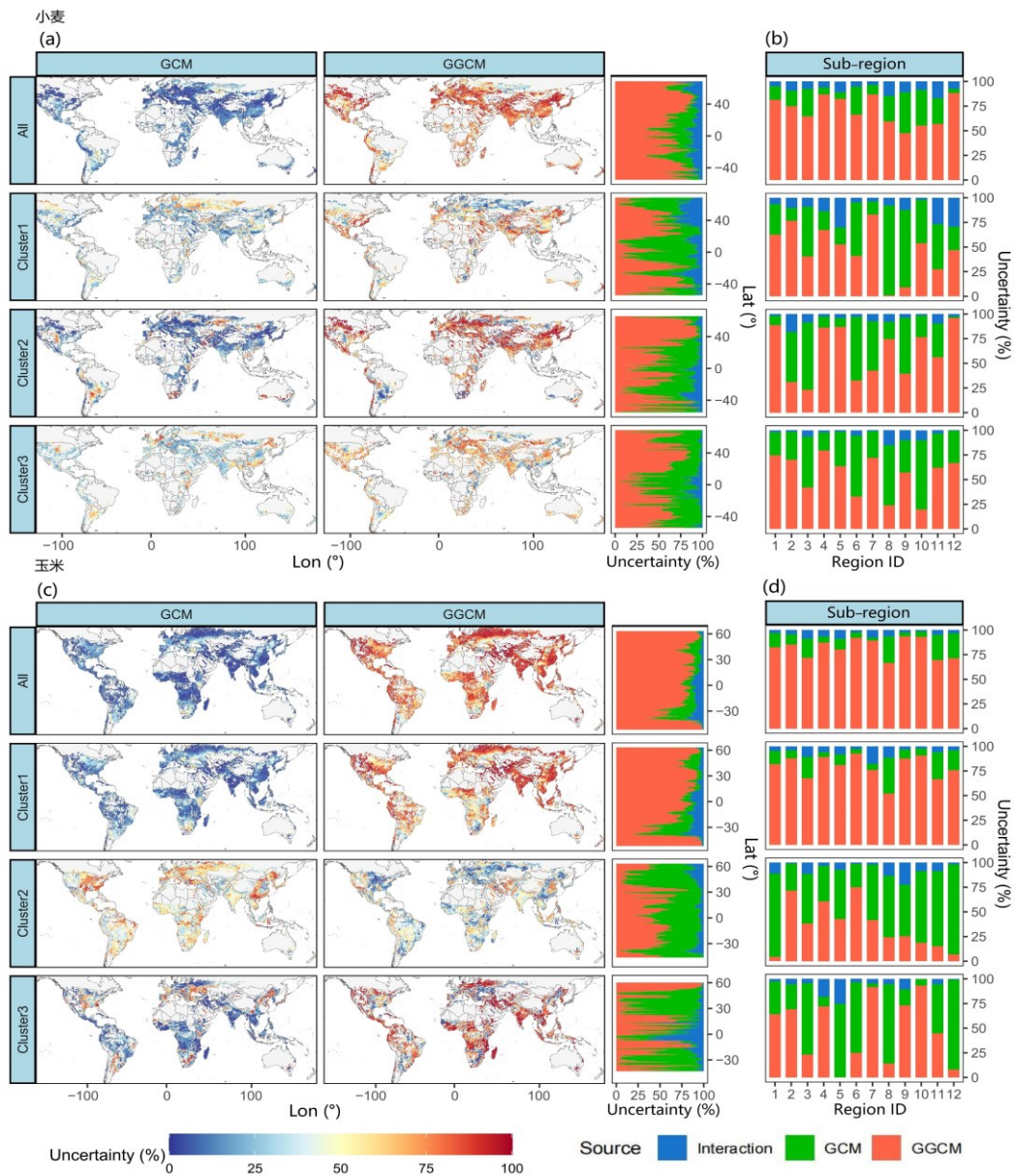


图 4-12 与 1980-2010 年相比，SSP585 在 2069-2099 年模拟小麦和玉米产量变化中 GCM 和 GGCM 的不确定性比例。a 和 c 的左侧面板是全球不同集群(图 4-4)和小麦和玉米的所有集合成员在网格尺度上来自 GCM 和 GGCM 的不确定比例的空间分布。右边的 a 和 c 显示了小麦和玉米在不同纬度的不确定性比例。b 和 d 显示了不同子区域不确定性来源

Fig. 4-12 The proportion of uncertainty from GCM and GGCM in simulated wheat and maize yield changes in 2069–2099 under SSP585 compared with 1980–2010. The left panels of **a** and **c** are the spatial patterns of uncertainty proportions from GCMs and GGCMs at the gridded scale across the globe for different clusters (see Figure 4-4) and all ensemble members for wheat and maize. The right panels of **a** and **c** show the proportion of uncertainty at different latitudes for wheat and maize. **b** and **d** show the sources of uncertainty at different sub-regions

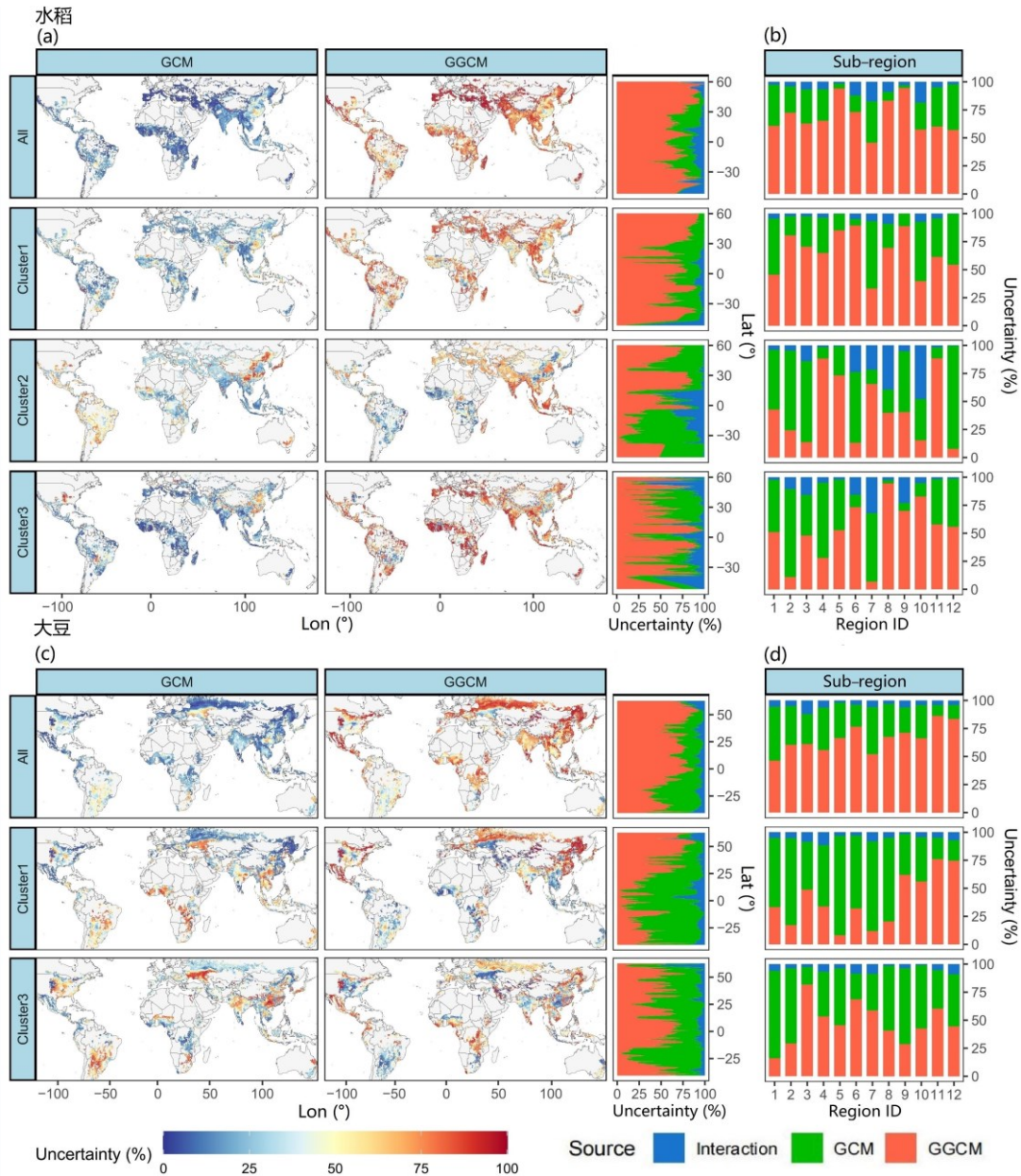


图 4-13 与 1980-2010 年相比, SSP585 在 2069-2099 年模拟水稻和大豆产量变化中 GCM 和 GGCM 的不确定性比例。a 和 c 的左侧面板是全球不同集群(图 4-5)和小麦和玉米的所有集成员在网格尺度上来自 GCM 和 GGCM 的不确定比例的空间分布。右边的 a 和 c 显示了小麦和玉米在不同纬度的不确定性比例。b 和 d 显示了不同子区域不确定性来源

Fig. 4-13 The proportion of uncertainty from GCM and GGCM in simulated rice and soybean yield changes in 2069–2099 under SSP585 compared with 1980–2010. The left panels of a and c are the spatial patterns of uncertainty proportions from GCMs and GGCMs at the gridded scale across the globe for different clusters (see Figure 4-5) and all ensemble members for wheat and maize. The right panels of a and c show the proportion of uncertainty at different latitudes for wheat and maize. b and d show the sources of uncertainty at different sub-regions

此外，不确定性的主要来源可能因地区而异。例如，GGCM 是北美（子区域 1 和 4）小麦产量变化的主要不确定性来源（集群 1-3 和所有模型集合），而 GCM 是南美（子区域 3）的主要不确定来源（集群 1-3）（图 4-12a 和图 4-13b）。

### 4.2.3 模型数量对不确定性的影响

为了减少模拟数量，一些研究对模型性能进行了排名，并选择了代表性模型进行进一步分析（Li et al. 2019a; Rattalino Edreira et al. 2021; Wang et al. 2016a），然而，所选模型数目可能无法反映整个系综的总体特征。在这里，量化了与模型数量相关的不确定性，包括九个 GGCM 和 32 个 GCM（图 4-14 和图 4-15）。结果表明，使用的模型数量（GGCM 或 GCM）越多，GGCM 或 GCM 引起的不确定性比例越高，但当模型数量达到一定数目时，不确定性变得稳定。本研究使用 $\pm 2.5\%$ （GGCM）和 $\pm 1.5\%$ （GCM）的阈值，从所选 GGCM 和 GCM 与全集合相比的不确定性贡献中得出，以确定可最小的模型数量（图 4-14 和图 4-15）。值得注意的是，在本研究中定义了采样不确定性所需的最少 GGCM 数量为当 GGCM 引起的不确定性在 GGCM 完整集合不确定性的 $\pm 2.5\%$ 以内时（Wang et al. 2020c）。然而，由于 GCM 引起的不确定性比例低于 GGCM，使用 $\pm 1.5\%$ 的阈值来选择 GCM 的最小集合大小。

研究发现，随机选择至少 4-6 个作物模型（包括小麦、玉米和水稻的 6 个 GGCM，以及大豆的 5 个 GGCM）和 9-12 个 GCM（包括小麦的 9 个 GCM，玉米的 10 个 GCM，水稻和大豆的 12 个 GCM），可以更好地捕捉产量变化的不确定性，与使用完整模型集相比具有相当的效果。

然而，本研究也遇到了一个问题，即是否可以通过使用特定的集成策略来进一步减小最小有效集合的大小。为解决这个问题，本研究采用了基于聚类的系统发育树来选择 GGCM 的子集，确保每个集群至少包含一个作物模型。如图 4-14 中阴影部分所示，筛选出 3-4 个模型。结果表明，使用 3-4 个 GGCM 即可捕捉 GGCM 引起的总体方差，这进一步说明了考虑模型类型对于设计多模型集合的重要性。

本研究提出的方法具有提高预测效率、降低冗余的重要意义。首先，确定最低的模型数目，可以在使用多个作物模型时减少冗余、提高预测效率，并保持较高的预测准确性。其次，基于聚类分析的方法进一步优化模型数目，可以在数据量较大时快速筛选出最具代表性的模型，提高预测效率。此外，优化模型数目还可以避免过多模型对预测结果的干扰，提高预测准确性，为作物种植者、政策制定者以及粮食生产者提供更准确、更有价值的信息。因此，本研究的方法具有高度的实用性和重要性。



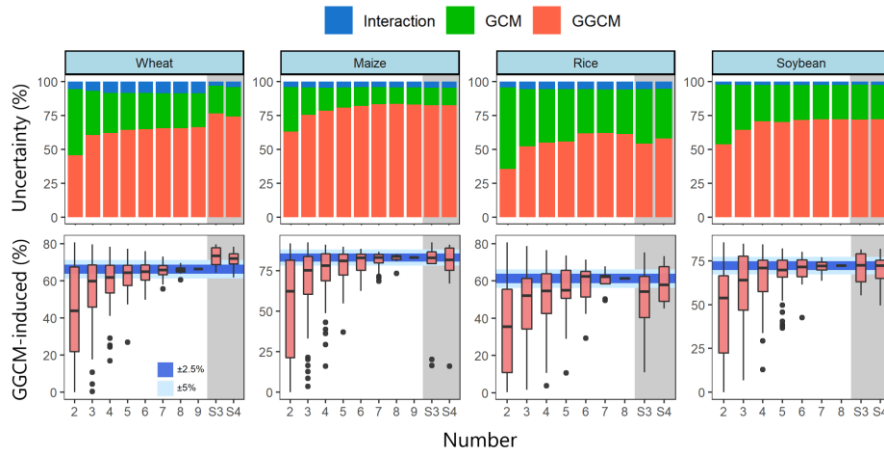


图 4-14 在 SSP585 下，小麦、玉米、水稻和大豆使用不同数量的 GGCM 时，产量变化的不确定性比例。数字 2-9(和 2-8)表示使用的 GGCM 的不同数量，包括小麦和玉米(水稻和大豆)的所有可能组合。箱线图显示了在使用不同数量的 GGCM 下，由于 GGCM 选择的不确定性的比例

Fig. 4-14 The proportion of uncertainty of yield changes (2069–2099 compared with 1980–2010, under SSP585) for different numbers of GGCMs used for wheat, maize, rice, and soybean. The numbers 2–9 (and 2–8) represent different numbers of GGCMs used, including all of the possible combinations for wheat and maize (rice and soybean). The boxplot shows the distribution of uncertainty due to GGCM-selection under different numbers of GGCMs used.

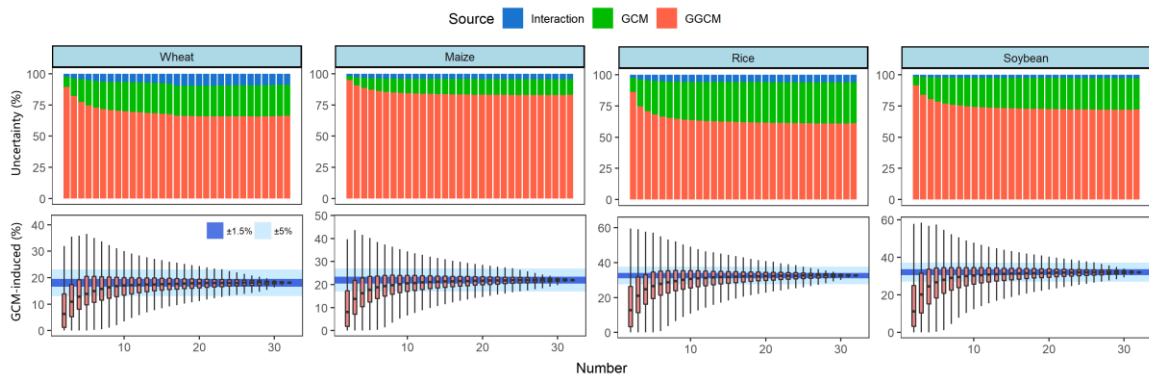


图 4-15 在 SSP585 下，小麦、玉米、水稻和大豆使用不同数量的 GCM 时，产量变化的不确定性比例。数字 2-32 表示使用的 GCM 的不同数量，包括小麦和玉米(水稻和大豆)的所有可能组合。

箱线图显示了在使用不同数量的 GCM 下，由于 GCM 选择的不确定性的比例

Fig. 4-15 The proportion of uncertainty of yield changes (2069–2099 compared with 1980–2010, under SSP585) for different numbers of GCMs used for wheat, maize, rice, and soybean. The numbers 2–32 represent different numbers of GCMs used, including all of the possible combinations for wheat and maize (rice and soybean). The boxplot shows the distribution of uncertainty due to GCM-selection under different numbers of GCMs used.



对于各个子区域来说,最小的有效集合数目有所不同。整体上作物模型数目在 4-6 左右,气候模式在 6-14 个左右(图 4-16 至 4-19 和表 4-2 至 4-3)。值得注意的是,本研究使用的分区是基于数据划分的,具有可以更充分的考虑不同区域的情况。目前,也有几个类似的区域化划分可用,包括 IPCC 的气候区 (Iturbide et al. 2020)、全球农业生态区 (AEZ) 和全球粮食生产单位 (FPU)。尽管存在这些计划,但这些计划可能无法有效捕捉作物产量不确定性来源的区域多样性。这是因为收获面积分布和影响作物产量的因素存在差异。

表 4-2 全球和区域尺度不确定性分析中 GGCM 的最小有效数目

Table 4-2 The minimum effective number of GGCM in uncertainty analysis at global and regional scales

区域	小麦	玉米	水稻	大豆
全球	6	6	6	5
区域 1	3	5	5	5
区域 2	5	6	5	5
区域 3	6	6	6	6
区域 4	4	5	3	5
区域 5	4	5	5	4
区域 6	5	5	6	4
区域 7	5	5	5	5
区域 8	6	6	5	5
区域 9	6	3	4	4
区域 10	8	4	5	5
区域 11	6	5	3	4
区域 12	3	6	5	5



图 4-16 在 SSP585 下，12 个子区域小麦、玉米、水稻和大豆使用不同数量的 GGCM 时，产量变化的不确定性比例。数字 2-9(和 2-8)表示使用的 GGCM 的不同数量，包括小麦和玉米(水稻和大豆)的所有可能组合

Fig. 4-16 The proportion of uncertainty of yield changes (2069–2099 compared with 1980–2010, under SSP585) for different numbers of GGCMs used for wheat, maize, rice, and soybean in 12 sub-regions. The numbers 2–9 (and 2–8) represent different numbers of GGCMs used, including all of the possible combinations for wheat and maize (rice and soybean)

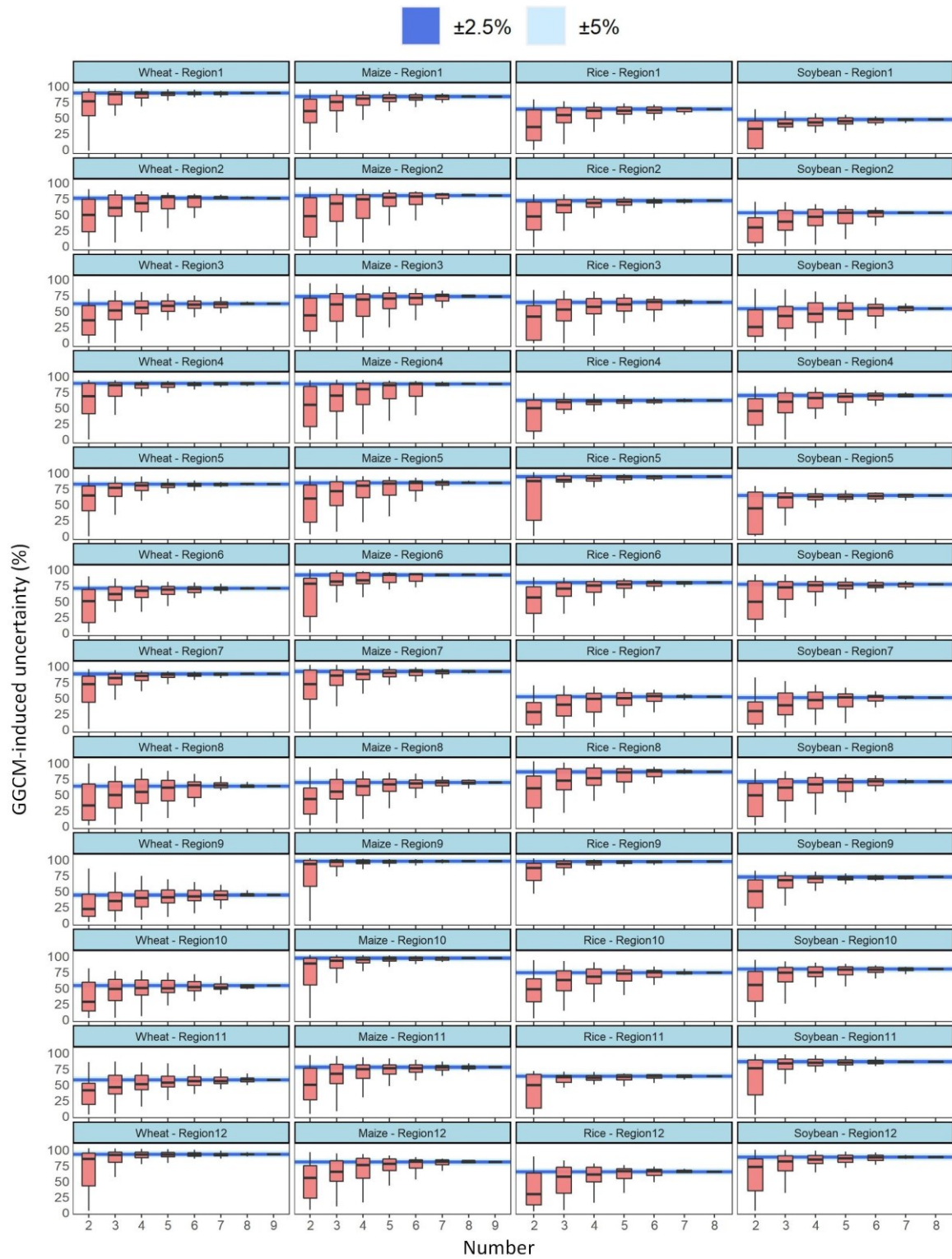


图 4-17 不同 GCM 数量下 GCM 贡献的不确定性的分布。

Fig. 4-17 The distribution of uncertainty due to GCM under different numbers of GGCMs used.



图 4-18 在 SSP585 下，12 个子区域小麦、玉米、水稻和大豆使用不同数量的 GCM 时，产量变化的不确定性比例。数字 2-32 表示使用的 GCM 的不同数量，包括四种作物的所有可能组合。

Fig. 4-18 The proportion of uncertainty of yield changes (2069–2099 compared with 1980–2010, under SSP585) for different numbers of GCMs used for wheat, maize, rice, and soybean in 12 sub-regions.

The numbers 2-32 represent different numbers of GCMs used, including all of the possible combinations for different crops.

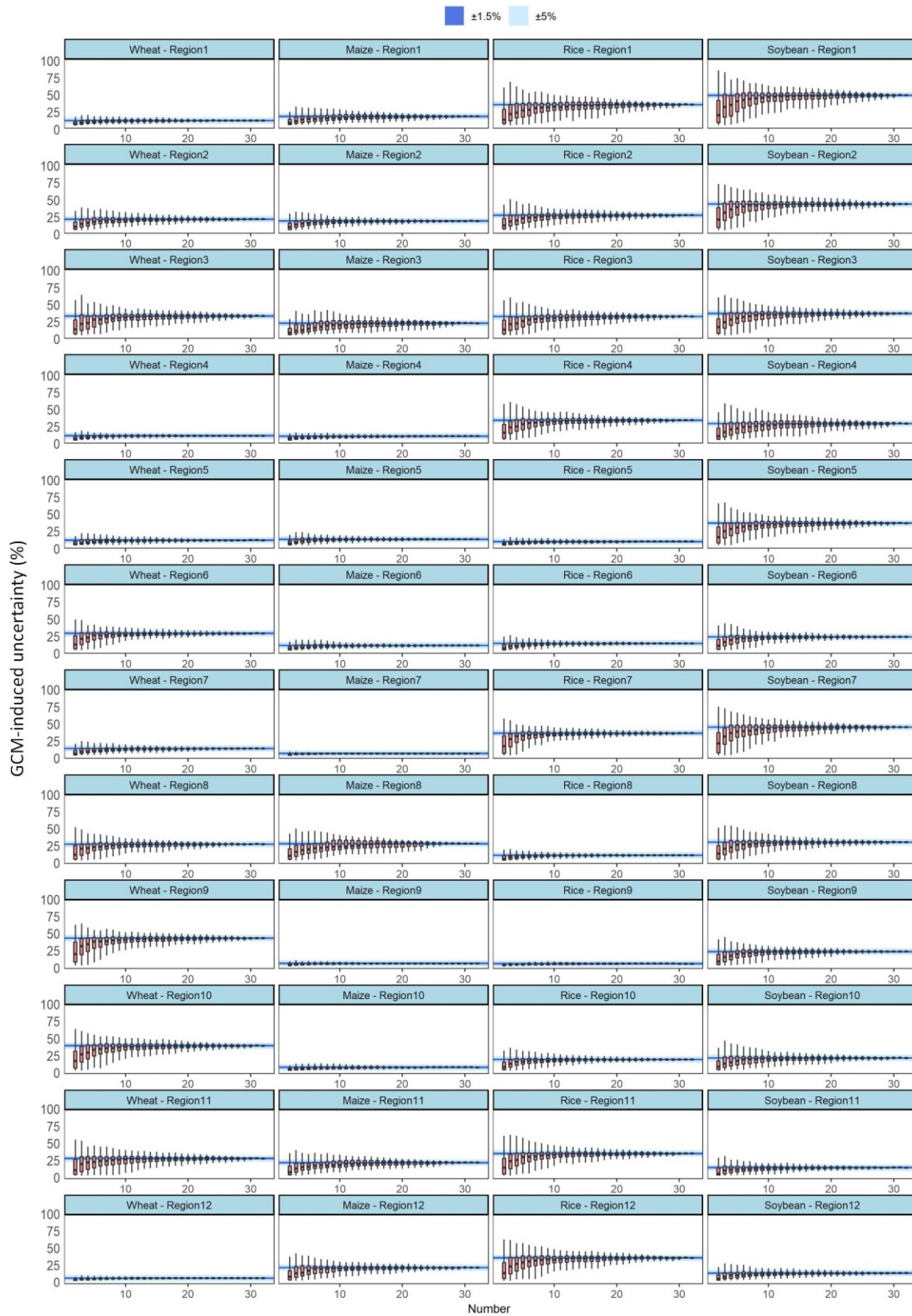


图 4-19 不同 GCM 数量下 GCM 贡献的不确定性的分布

Fig. 4-19 The distribution of uncertainty due to GCM under different numbers of GCMs used

表 4-3 全球和区域尺度不确定性分析中 GCM 的最小有效数目

Table 4-3 The minimum effective number of GCM in uncertainty analysis at global and regional scales

区域	小麦	玉米	水稻	大豆
全球	9	10	12	12
区域 1	6	11	14	13
区域 2	10	8	11	12
区域 3	12	13	14	13
区域 4	5	6	13	13
区域 5	6	6	6	13
区域 6	10	6	7	10
区域 7	8	3	11	13
区域 8	11	15	6	12
区域 9	13	3	4	10
区域 10	12	6	9	7
区域 11	11	13	13	6
区域 12	4	7	12	7

此外，当模型数量大于 5 个时，GGCM 导致的不确定性和相关标准偏差的空间分布上变化不大(图 4-20 和图 4-21)。这对于选择足够数量的气候和作物模型来模拟气候变化下的产量响应是至关重要的信息。



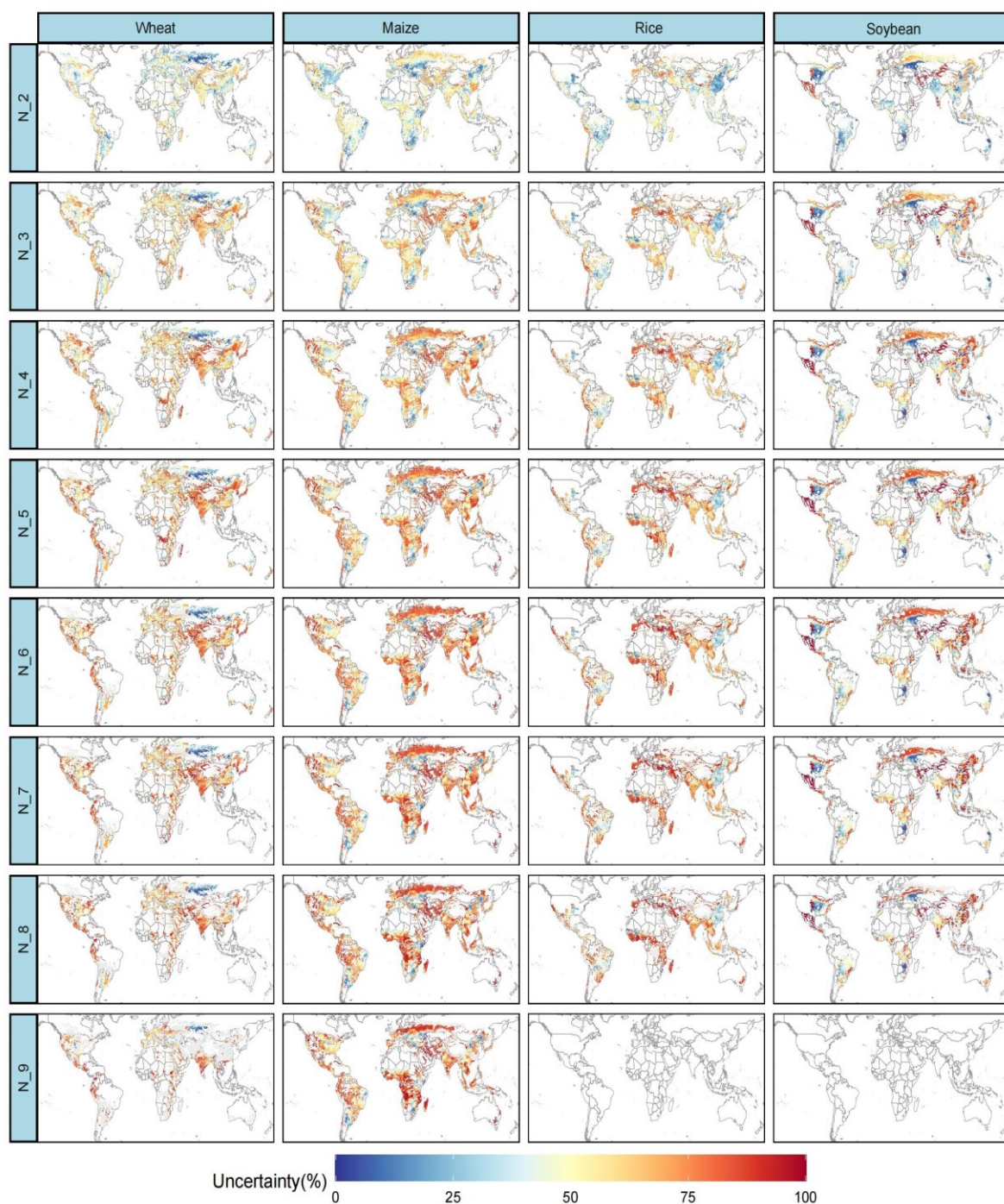


图 4-20 不同 GGCM 数量组合下 GGCM 诱导作物产量不确定性的空间格局, 其中 N\_2-N\_9 表示 GGCM 的数量(从 2 到 9), 由于水稻和大豆有 8 个 GGCM, N\_9 没有水稻和大豆的数据

Fig. 4-20 Spatial patterns of GGCM-induced crop yield uncertainty under combinations of different numbers of GGCMs. N\_2-N\_9 represent the number of GGCMs (from 2 to 9). No data for N\_9 as there are 8 GGCMs available for rice and soybean



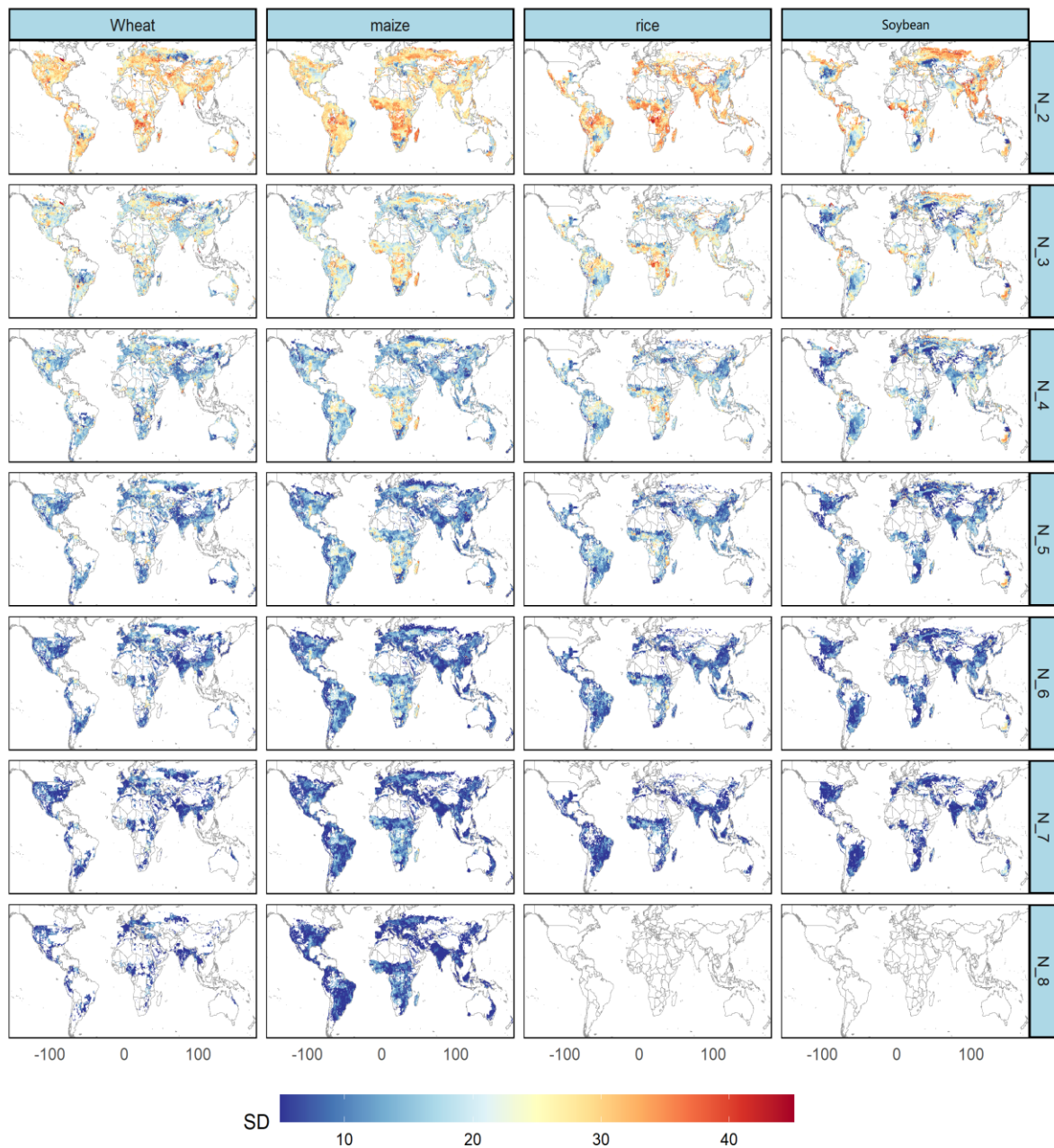


图 4-21 不同数量 GGCMd 贡献的作物产量不确定性的标准差(SD)。例如, N\_2 是从 9 个(小麦和玉米)或 8 个(水稻和大豆)GGCM 中随机选取的 2 个。

Fig. 4-21 The standard deviation (SD) of GGCM-induced crop yield uncertainty when different numbers of GGCMs are used. For example, N\_2 is two GGCMs are randomly selected from 9 (wheat and maize) or 8 (rice and soybean) GGCMs.

为了比较第 3 阶段改进后的 GGCMi 数据集中模型的效果, 本章确定了最小有效的集成大小并进行了分析(详见图 4-22)。研究发现, 随着 GGCM 数量的增加, 模拟的不确定性模式与第 2 阶段相似, 并且使用六到七个 GGCM (小麦和大豆的七个

GGCM, 玉米和水稻的六个 GGCM) 可以反映产量变化预测的总体不确定性。由于第 3 阶段只有五个 GCM, 没有分析对其总体不确定性的影响。通过基于聚类分析选择模型, 可以将集成大小进一步减少至 3-4 个模型, 从而反映总体不确定性。然而, 对于小麦产量预测来说, 情况可能并非如此 (详见图 4-22)。小麦产量的聚类树 (GGCM 阶段 3) 并不总是以 GGCM 为中心, 如图 4-6 中的聚类 1 所示, 该聚类主要由 IPSL-CM6A-LR 和 UKESM1-0-LL 分组。尽管结果表明使用 GGCM 的子集足以捕获总体方差, 但在直接使用多模型集合投影时仍应保持谨慎。

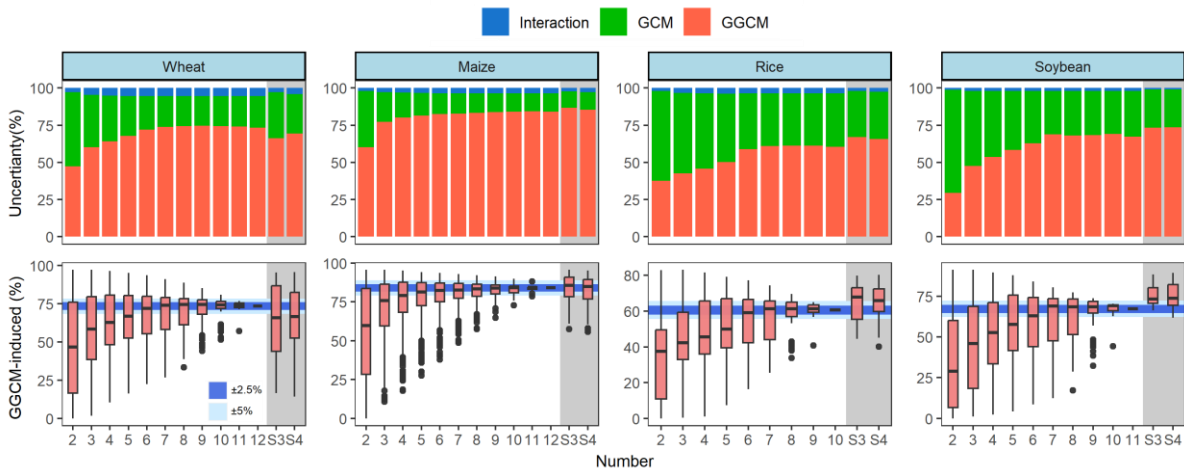


图 4-22 在 SSP585 下, 小麦、玉米、水稻和大豆使用不同数量的 GGCM (第三阶段) 时, 产量变化的不确定性比例。数字 2-12 表示使用的 GGCM 的不同数量, 包括不同作物的所有可能组合。

箱线图显示了在使用不同数量的 GGCM 下, 由于 GGCM 选择的不确定性的比例

Fig. 4-22 The proportion of uncertainty of yield changes (2069–2099 compared with 1980–2010, under SSP585) for different numbers of GGCMs (GGCMI phase 3) used for wheat, maize, rice, and soybean.

The numbers 2-12 represent different numbers of GGCMs used, including all of the possible combinations for different crops. The boxplot shows the distribution of uncertainty due to GGCM-selection under different numbers of GGCMs used.

#### 4.2.4 单一模型对整体不确定性的影响

在许多地区, 全球气候模型 (GGCM) 是主要的不确定性来源。因此, 本章重点研究归因于单个 GGCM 的不确定性。本研究通过删除每个作物模型并比较完整模型组和不使用特定模型的 GGCM 引起的不确定性的差异, 测试了单个模型引起的不确定性比例 (见图 4-23 至图 4-26)。对于小麦而言, CARAIB、PEPIC 和 pDSSAT 模型在大多数地区增加了不确定性, 而 LPJ-GUESS 和 JULES 的贡献则因地区而异 (见图 4-23)。例如, LPJ-GUESS 模型增加了澳大利亚小麦产量预测中 GGCM 引起的不确定性份额, 但降低了东南亚 GGCM 引起的不确定性份额。这种可变结果可能是因为 LPJ-GUESS 模型通常在强 CO<sub>2</sub> 施肥反应中表现异常 (Franke et al. 2020a)。

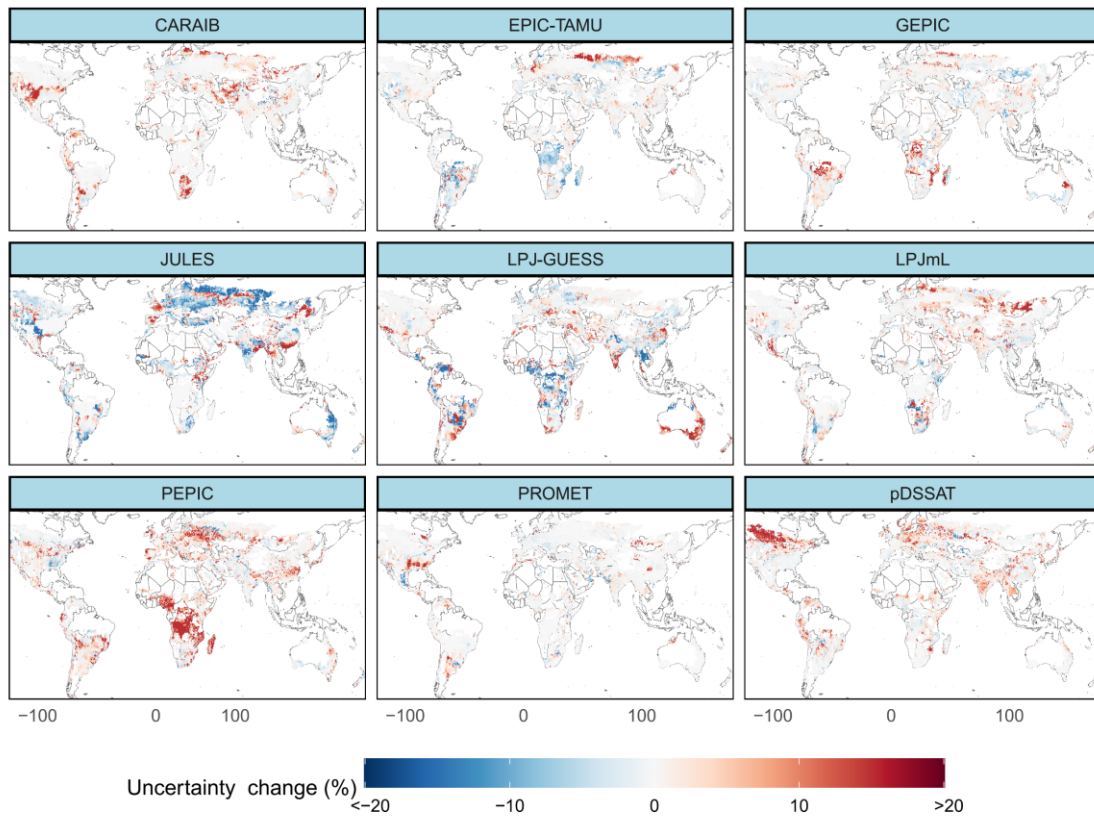


图 4-23 在 SSP585 下, 2069–2099 年, 各个 GGCM 在网格级别对小麦造成的 GGCM 引起的不确定性份额的变化。通过在每个网格点从 GGCM 集合中依次删除九个 GGCM 中的一个, 然后比较完整 GGCM 模型集合和没有特定模型的集合之间 GGCM 引起的预计产量变化的不确定性份额, 来确定贡献。积极的变化表明, 包括特定的模型增加了 GGCM 引起的不确定性份额。负贡献变化表明, 使用该模型降低了 GGCM 引起的不确定性份额

Fig. 4-23 Changes in the GGCM-induced uncertainty share contributed by individual GGCMs at the grid level for wheat in 2069–2099 under SSP585. The contribution was determined by sequentially removing one of the nine GGCMs from the ensemble of GGCMs at each grid point, and then comparing the GGCM-induced uncertainty share of projected yield changes between the full GGCM model set and the set without the particular model. A positive change indicates that including a particular model increased the GGCM-induced uncertainty share. A negative contribution change indicates that using this model decreased the GGCM-induced uncertainty share. GGCM is global gridded crop model

对玉米来说, 可以看到 PEPIC 和 pDSSAT 增加了 GGCM 诱导的玉米产量预测不确定性份额。这可能是由于 pDSSAT 对玉米具有强烈的温度响应(Franke et al. 2020a)。JULES 增加了南半球 GGCM 引起的不确定性份额, 并减少了主要在北半球 GGCM 引起的不确定性(见图 4-24)。就水稻而言, PEPIC 主要促成了大多数地区, 特别是非洲地区 GGCM 引起的不确定性; JULES 增加了 GGCM 在亚洲引起的不确定性份额;



而 PROMET 和 pDSSAT 对 GGCM 诱导的不确定性份额产生了很大的影响，但方向在不同的区域是不同的（见图 4-25）。对于大豆，JULES 增加了几乎所有地区 GGCM 引起的不确定性份额；CARAIB、PROMET 和 pDSSAT 也对 GGCM 诱导的不确定性份额产生了很大的影响，但这种影响在不同的区域之间也是不同的（见图 4-26）。这些结果表明，每个 GGCM 都在一定程度上影响了全球模型预测的不确定性，突显了在研究区内选择模型的重要性

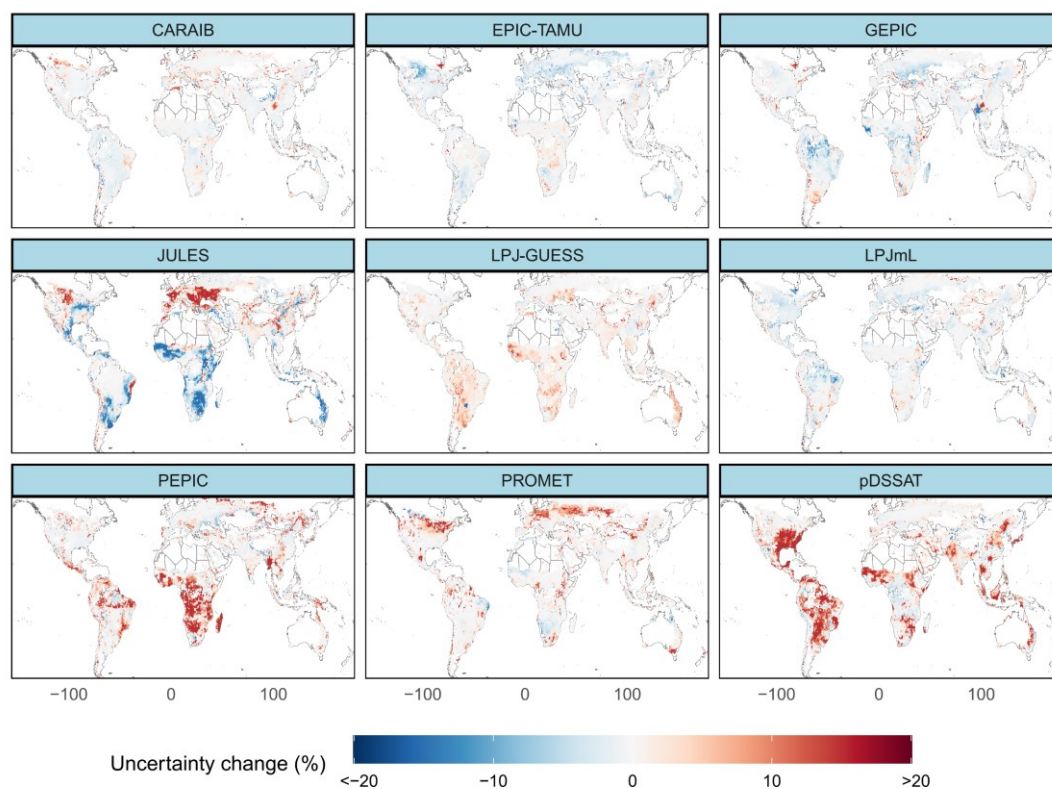


图 4-24 在 SSP585 下，2069–2099 年，各个 GGCM 在网格级别对玉米造成的 GGCM 引起的不确定性份额的变化。通过在每个网格点从 GGCM 集合中依次删除九个 GGCM 中的一个，然后比较完整 GGCM 模型集合和没有特定模型的集合之间 GGCM 引起的预计产量变化的不确定性份额，来确定贡献。积极的变化表明，包括特定的模型增加了 GGCM 引起的不确定性份额。负贡献变化表明，使用该模型降低了 GGCM 引起的不确定性份额

Fig. 4-24 Changes in the GGCM-induced uncertainty share contributed by individual GGCMs at the grid level for maize in 2069–2099 under SSP585. The contribution was determined by sequentially removing one of the nine GGCMs from the ensemble of GGCMs at each grid point, and then comparing the GGCM-induced uncertainty share of projected yield changes between the full GGCM model set and the set without the particular model. A positive change indicates that including a particular model increased the GGCM-induced uncertainty share. A negative contribution change indicates that using this model decreased the GGCM-induced uncertainty share. GGCM is global gridded crop model

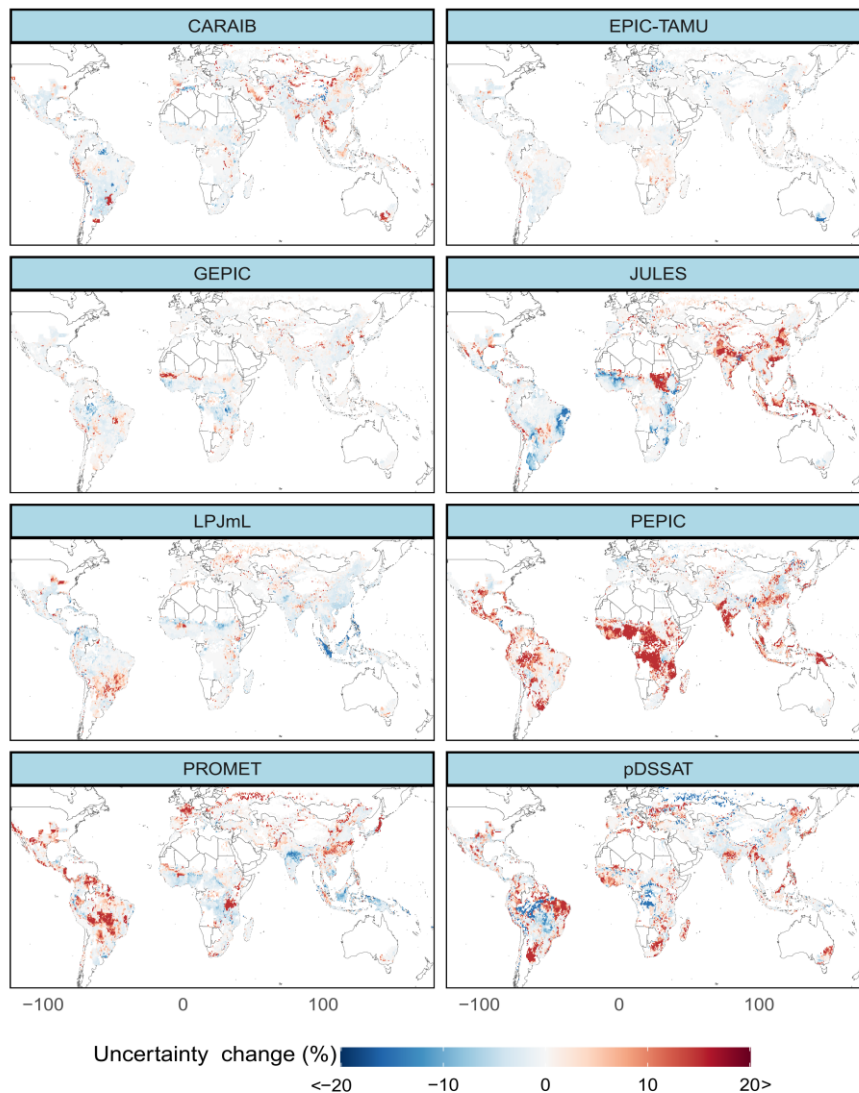


图 4-25 在 SSP585 下，2069–2099 年，各个 GGCM 在网格级别对水稻造成的 GGCM 引起的不确定性份额的变化。通过在每个网格点从 GGCM 集合中依次删除八个 GGCM 中的一个，然后比较完整 GGCM 模型集合和没有特定模型的集合之间 GGCM 引起的预计产量变化的不确定性份额，来确定贡献。积极的变化表明，包括特定的模型增加了 GGCM 引起的不确定性份额。负贡献变化表明，使用该模型降低了 GGCM 引起的不确定性份额

Fig. 4-25 Changes in the GGCM-induced uncertainty share contributed by individual GGCMs at the grid level for rice in 2069–2099 under SSP585. The contribution was determined by sequentially removing one of the eight GGCMs from the ensemble of GGCMs at each grid point, and then comparing the GGCM-induced uncertainty share of projected yield changes between the full GGCM model set and the set without the particular model. A positive change indicates that including a particular model increased the GGCM-induced uncertainty share. A negative contribution change indicates that using this model decreased the GGCM-induced uncertainty share. GGCM is global gridded crop model

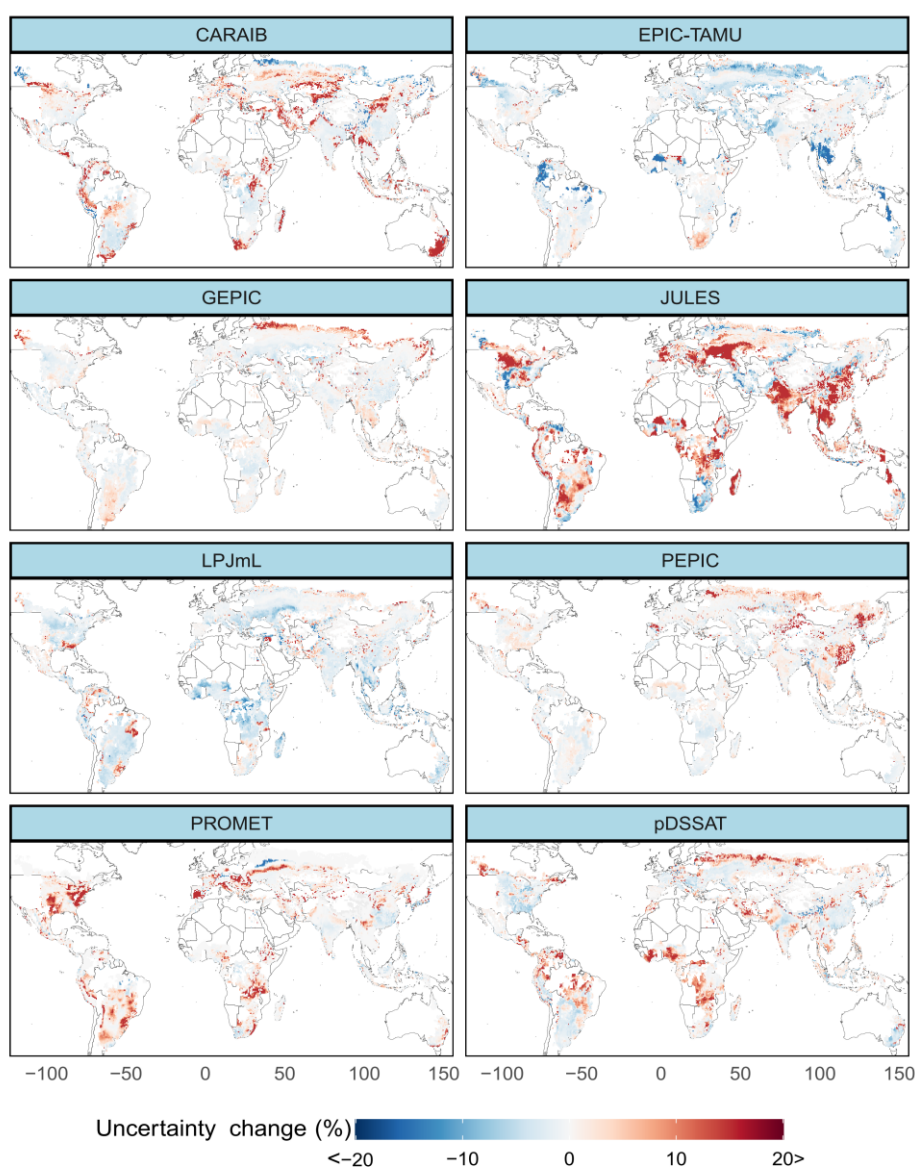


图 4-26 在 SSP585 下，2069–2099 年，各个 GGCM 在网格级别对水稻造成的 GGCM 引起的不确定性份额的变化。通过在每个网格点从 GGCM 集合中依次删除八个 GGCM 中的一个，然后比较完整 GGCM 模型集合和没有特定模型的集合之间 GGCM 引起的预计产量变化的不确定性份额，来确定贡献。积极的变化表明，包括特定的模型增加了 GGCM 引起的不确定性份额。负贡献变化表明，使用该模型降低了 GGCM 引起的不确定性份额

Fig. 4-26 Changes in the GGCM-induced uncertainty share contributed by individual GGCMs at the grid level for rice in 2069–2099 under SSP585. The contribution was determined by sequentially removing one of the eight GGCMs from the ensemble of GGCMs at each grid point, and then comparing the GGCM-induced uncertainty share of projected yield changes between the full GGCM model set and the set without the particular model. A positive change indicates that including a particular model increased the GGCM-induced uncertainty share. A negative contribution change indicates that using this model decreased the GGCM-induced uncertainty share. GGCM is global gridded crop model

## 4.3 讨论

### 4.3.1 预测的不确定性以及模型选择

尽管多模型集合平均值是预测作物产量的常用统计数据，但它在确定适应策略方面的效用受到限制。这些策略需要考虑一系列可能的未来产量轨迹，并且当单个模型组合中存在不同符号的变化时，平均值可能会导致误导性结果。因此，需要考虑不同模型集合的结果。研究结果可以提供有用的信息去用来更好的设计 GCM-GGCM 集合，以最大化不确定性采样。本研究研究聚焦于 SSP585 排放轨迹，以探索大气候变化信号下的产量变化。Müller 等 (2021) 发现，对于这些 GGCM 仿真器，SSP126 和 SSP585 之间的产量预测中不同不确定性来源的相对重要性是具有可比性的，在 SSP585 下，总体交叉集合方差显著较大。因此，结果相比于其它情景具有代表性。

本章分析了集合大小对不确定性的影响。本研究发现，模型的数量对全球不同地区的不确定性的影响有很大差别。例如，使用 10 个 GCM 捕获了全球范围内 GCM 引起的玉米产量变化的不确定性 (图 4-19)，但美国、中亚和南美洲南部 (分区 1、3、8 和 11) 需要更多的 GCM (图 4-18 和表 4-4)。

不同集合成员的不同产量变化对多模型集合的设计具有重要意义 (图 4-4 和图 4-5)。例如，在南亚和中欧，JULES 对大豆产量的模拟与其他模型的结果不同，对总体不确定性做出了很大贡献。如何处理这种极端行为可能会对多模型集合的结果和相应的不确定性产生很大影响 (Knutti 2010)。在这些情况下，一些先前的研究将给定的模型排除在外或降低了权重，以减少总体不确定性 (Vetter et al. 2017; Wang et al. 2020b)。因此，在选择用于多模型集成以减少气候变化影响评估中的不确定性之前，仔细筛选和评估各个模型的性能，以确保其稳健性，从而模拟气候变化对作物产量的影响是非常重要的 (Wang et al. 2017a)。

多模型集合在预测作物产量变化时具有一定的局限性，因为不同模型之间的结果可能会相互抵消，导致预测结果相对保守。然而，本研究通过聚类分析方法，针对模型集合中的多个模型进行分类，避免了不同模型之间的结果相互抵消的情况。通过聚类分析，可以将相似的模型分为同一类别，同时排除过多模型带来的冗余，从而提高了预测效率和准确性。因此，本研究采用聚类分析对模型选择具有重要意义，能够优化模型集合，避免结果的抵消过程，提高预测准确性，为制定适应策略和维持可持续性提供更为可靠的依据。

### 4.3.2 不同的不确定性来源

目前已经开发了不同的方法来提高作物模型的性能并降低整体建模的不确定性，例如改进模型结构 (Maiorano et al. 2017) 或参数优化 (Wallach et al. 2021; Xiong et al. 2019)。然而，由于高计算成本和缺少校准目标数据，这项工作在全球范围内实施具



有挑战性。最近，一些研究使用田间实验数据和统计模型的结果对 GGCM 的不确定性进行约束，提出了使用更详细的野外实验数据可以改善模型对气候变化的响应，并进一步降低总体不确定性 (Wang et al. 2020e; Zhao et al. 2016b)。除此之外，二氧化碳施肥和其他因素也可能对作物生长产生影响，从而增加产量预测的不确定性。例如，几项研究发现，CO<sub>2</sub> 施肥效应受到频繁的极端天气事件的制约 (Gray et al. 2016; Obermeier et al. 2016)，但这些影响目前的模拟并未充分考虑，可能导致产量损失的低估和总体不确定性的增加 (Wang et al. 2020e)。因此，为了更好地理解气候变化对作物产量的影响，减少气候变化影响评估中的不确定性，需要在更大范围、更长时间、更大规模的作物基因型受控环境实验中研究气候变化 (如热浪、极端降水、干旱和 CO<sub>2</sub>) 与作物产量之间的相互影响 (Kimball 2016; Ottman et al. 2012)。

### 4.3.3 本研究的局限性以及展望

本研究中有几个局限性。一些局限性于 GGCM 模拟有关。首先，GGCM 模拟没有考虑土壤数据和管理措施 (例如施肥率) 等不确定性来源 (Folberth et al. 2016; Folberth et al. 2019; Hasegawa et al. 2021; McCullough et al. 2022) 以及病虫害的影响 (Rosenzweig et al. 2014a)。其次，模型参数化也会影响模拟并影响不确定性。例如，Xiong 等 (2019) 使用四种参数化策略分析了参数化产生的不确定性。虽然本研究中考虑了 GGCM 之间的结构差异，但没有考虑单个 GGCM 参数设置的不确定性。第三，尽管一些 GGCM 可以捕捉极端温度事件 (包括高温和霜冻) 和干旱情况下发生的产量损失 (Schauberger et al. 2017a)，但它们低估了极端潮湿条件下的产量损失程度 (Heinicke et al. 2022)。最后，使用的变化因子方法来生成未来气候的数据也存在不确定性。与 GCM 模拟相似，该方法没有充分采样极端气候事件的变化，也没有考虑与季节内变化相关的事件变化 (模拟器仅由作物生长季节的气候变化驱动)，从而导致不确定性。虽然本研究使用了 GGCM 第 3 阶段的数据集，该数据集使用偏差调整的每日 GCM 数据与 GGCM 模拟器进行比较，主要考虑了不确定性来源，但目前只有五个 GCM 可用，无法充分反映 GCM 造成的方差。因此，需要更多的努力，以更好地协调气候和作物模型的过程，以期望从不同的模型集合中获得接近“确定”的预测。

## 4.4 小结

在本研究中，不同气候条件下，作物模型不确定性的来源存在区域差异，这主要取决于使用的气候和作物模型。为了评估气候变化对农业的影响，研究人员通常使用关联的气候和作物模拟模型，但模型组合和数量对不确定性的影响尚未得到明确的研究。因此，研究了全球网格作物模型 (GGCM) 和全球气候模型 (GCM) 在模拟未来气候下，全球作物产量变化方面的集合配置对建模不确定性的影响。本研究发现，

在不同 GCM 和 GGCM 的组合中，气候变化下建模不确定性的主要来源存在较大的差异，尤其是对于玉米、水稻和大豆等作物，而小麦的不确定性相对较小。

此外，本研究随机选择大约 6 个 GGCM 和 10 个 GCM 来确定 9 个作物模型和 32 个气候模型的建模不确定性。然而，本研究进一步使用聚类分析的结果减少最小集合数目，发现在保证每一个聚类至少有一个模型的情况下，大约 3-4 个模型即可反映整体的方差。这些结果突出了模型组成和集合大小在确定作物产量预测中不确定性的主要来源方面的重要性。这项研究的发现为研究气候变化对农业产量的影响提供了有价值的信息和指导，并为农业政策制定者提供了有用的参考，帮助他们更好地应对未来的气候变化。未来的研究可以进一步探索其他因素（例如土壤数据和管理措施以及病虫害的影响）对不确定性的影响，并研究更多可能的模型组合和集合大小对建模不确定性的影响，以便为精确预测未来气候下的作物产量提供更准确的信息。

## 第五章 基于机器学习提高网格作物模型对产量的预测并约束不确定性

作物模型虽然可以较好的解释作物生长发育过程，但是作物模型在模拟中往往过于简化极端气候的过程（例如高温、低温、干旱、暴雨等）。此外，当前的作物模型的一个普遍的问题是，没有考虑病虫害风险对产量的影响。也有一些研究利用多模型集合预测气候变化对产量的影响。然而作物模型本身对于作物生长发育和产量的响应也存在差异，不同的作物模型对于环境因素的响应程度和机理不同，也会导致不同气候模式驱动的作物模型预测结果存在较大不确定性。而基于机器学习虽然可以得到较为可靠的产量模拟，但是由于机器学习只是简单的拟合缺乏对一些影响因素（如 CO<sub>2</sub> 施肥效应、管理措施等）的深入理解。因此，有必要构建混合模型以结合两者的优点，考虑极端气候等因素对产量影响，同时可以约束作物模型的不确定性。因此，本研究将以玉米和大豆为例（1）结合网格化作物模型以及机器学习考虑不同因素（极端气候与病虫害风险），构建混合模型；（2）分析历史和未来不同排放情景下气候变化对玉米和大豆产量的影响；（3）对比机器学习约束前后的不确定性变化以及区域差异。

### 5.1 材料和方法

#### 5.1.1 数据介绍

##### 5.1.1.1 气候数据

本研究选择了 21 个全球气候模式（GCM）从 1961 至 2100 年，的月尺度数据。从 CMIP6 官网获取（<https://esgf-node.llnl.gov/projects/cmip6/>）。其中包括 SSP126 和 SSP585 两个排放情景（表 5-1）。

表 5-1 21 个 CMIP6 气候模式的信息

Table 5-1 Information of the selected 21 GCM from CMIP6

编号	名称	缩写	分辨率 (°×°)
1	ACCESS-CM2	ACC1	1.2 × 1.8
2	ACCESS-ESM1-5	ACC2	1.2 × 1.8
3	BCC-CSM2-MR	BCCC	1.1 × 1.1
4	CanESM5	Can1	2.8 × 2.8
5	CIESM	CIES	0.9 × 1.3
6	CNRM-ESM2-1	CNR1	1.4 × 1.4
7	CNRM-CM6-1	CNR2	1.4 × 1.4
8	CNRM-CM6-1-HR	CNR3	1.4 × 1.4

表 5-1 (续)

编号	名称	缩写	分辨率 (°×°)
9	EC-Earth3-Veg	ECE2	0.7 × 0.7
10	FGOALS-g3	FGOA	5.2 × 2.0
11	GFDL-ESM4	GFD2	1.0 × 1.3
12	GISS-E2-1-G	GISS	2.0 × 2.5
13	INM-CM4-8	INM1	1.5 × 2.0
14	INM-CM5-0	INM2	1.5 × 2.0
15	MIROC6	MIR1	1.4 × 1.4
16	MIROC-ES2L	MIR2	2.7 × 2.8
17	MPI-ESM1-2-HR	MPI1	0.9 × 0.9
18	MPI-ESM1-2-LR	MPI2	1.8 × 1.9
19	MRI-ESM2-0	MTIE	1.1 × 1.1
20	NESM3	NESM	1.9 × 1.9
21	UKESM1-0-LL	UKES	1.3 × 1.9

## 5.1.1.2 田间实验数据

本研究从中国气象数据共享网络 (<http://data.cma.cn/>) 收集了来自 155 个试验点的玉米产量试验数据和 50 个试验点的大豆产量试验数据, 涵盖了 1999 年至 2010 年的信息, 包括作物产量和生长阶段的信息 (图 5-1)。

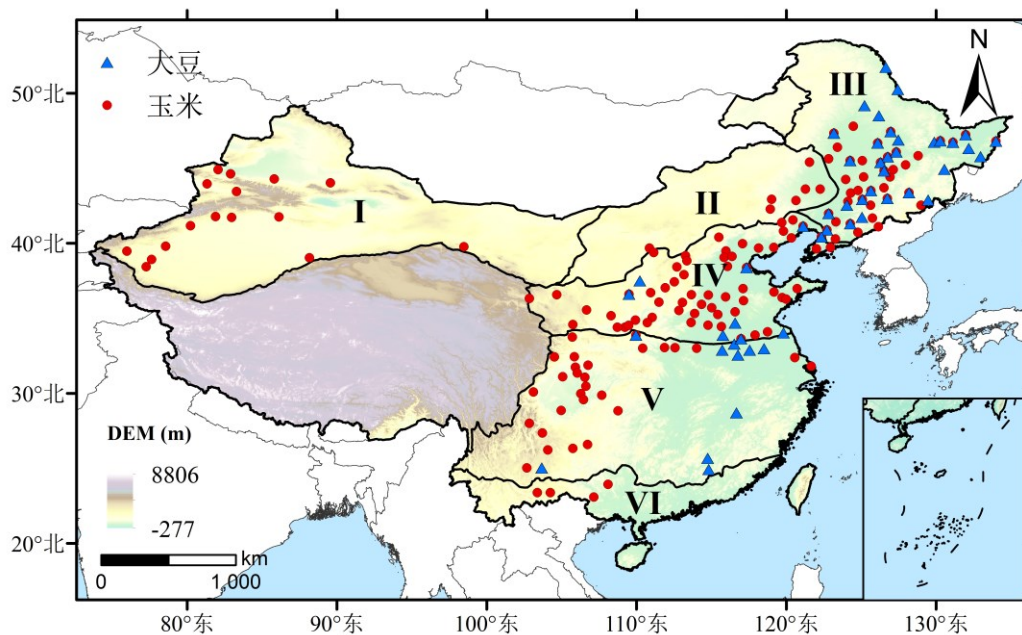


图 5-1 玉米大豆站点空间分布

Fig. 5-1 The spatial distribution of maize and soybean site in China

其中管理实践和收获方法与当地农民采用的方法一致。数据的管理实践和收获方法与当地农民的方法一致。然而，由于 1999 年至 2013 年期间观测到的作物产量数据不连续，因此共获得了 1076 组玉米田间数据和 368 组大豆田间数据。其中玉米包括了四个分区，大豆包括三个分区，分区方式见第三章。

### 5.1.2 NWAI-WG 统计降尺度方法

NWAI-WG 统计降尺度方法是基于天气发生器的一种降尺度方式。在空间上，选择站点所临近的四个网格进行反距离加权插值方法来插值到站点上。通过 QQ 图来对比观测与模拟的差异。基于概率分布函数来对模拟值进行校准。

由于本研究选择的气候模式原始数据为月尺度，因此需要在时间上做降尺度，对于降水来说主要基于伽马函数和一阶马尔可夫链来生成。其中，一阶马尔可夫链用于生成日序列，伽马分布用来计算降水概率。伽马分布的密度函数如下：

$$f(p) = \frac{p^{\alpha-1} e^{-\frac{p}{\beta}}}{\beta^{\alpha} \Gamma(\alpha)} \quad p, \beta > 0, 0 < \alpha < 1 \quad (5-1)$$

温度数据由 WGEN 天气发生器来生成，其计算方式与降水相似。利用序列相关系数与互相关系数：

$$X_i(j) = AX_{i-j}(j) + \varepsilon_i(j) \quad (5-2)$$

式中  $X(j)$  是包含第  $i$  天的三个气候变量的矩阵（包括最高温度，最低温度和辐射）。 $\varepsilon_i$  是独立的随机变量。A 和 B 是由下式定义的矩阵：

$$A = M_1 M_0^{-1}, \quad BB^T = M_0 - AM_0^T \quad (5-3)$$

式中  $M_0$  的元素是同一天这三个变量之间的相关系数， $M_1$  的元素是滞后 1 天的相关系数。

由于概率估算的降水有一定的随机性，因此日降水的数据总和总是不等于输入的月值。因此通过 1500 次循环使其差别达到一个可接受的范围。

该降尺度方法高效、灵活，并且有较好的偏差校准的效果，因此受到了广泛的应用。更详细的信息见 Liu and Zuo (2012)。

### 5.1.3 GGCM 模拟器

本研究使用了玉米和大豆中国区域的产量。玉米模拟包括了九个模型：CARAIB、EPIC-TAMU、JULES、GEPIC、LPJ-GUESS、LPJmL、pDSSAT、PEPIC 和 PROMET；大豆模拟包括了八种模型：CARAIB、EPIC-TAMU、JULES、GEPIC、LPJmL、pDSSAT、PEPIC 和 PROMET。更多信息见第四章。

由于一些 GGCM 在模拟的基准作物生产力水平上往往有所不同 (Müller et al.

2017), 对作物模型输出的产量做了校正以匹配观测期间的产量。具体方法如下:

$$Y_t^* = Y_{t,c} \times \frac{O_{r,c}}{Y_{r,c}} \quad (5-4)$$

其中  $Y_t^*$  是时间步长  $t$  的校准后的产量;  $Y_t$  是时间步长  $t$  的模拟产量;  $O_{r,c}$  是历史时段在  $c$  站点的产量 ( $r$  为时段, 本文为 1999-2010)。  $Y_{r,c}$  是历史时段该点的模拟值。

#### 5.1.4 不确定性分析

本文采用方差分析方法 (ANOVA) 对玉米和大豆的不确定性来源进行量化。本研究考虑了气候模式、作物模型和排放情景这三个因素作为主要的不确定性来源:

$$SA = SS_{GGCM} + SS_{GCM} + SS_{Scen} + SS_{GCM \times GGCM} + SS_{GCM \times Scen} + SS_{GGCM \times Scen} + SS_{GGCM \times GCM \times Scen} \quad (5-4)$$

其中  $SS_{GGCM}$ 、 $SS_{GCM}$  和  $SS_{Scen}$  代表由 GCM、GGCM 和排放情景带来的不确定性。 $SS_{GCM \times GGCM}$ 、 $SS_{GCM \times Scen}$ 、 $SS_{GGCM \times Scen}$  和  $SS_{GGCM \times GCM \times Scen}$  代表他们之间的交互影响。

#### 5.1.5 混合模型建模方式

由于作物模型对于极端气候相响应过程过于简化, 本章考虑了五个极端指数用于机器学习: SPEI: 主要用于反应干湿变化; 极端高温指数 (TD): 最高温大于  $30^\circ\text{C}$  的天数用于反应高温胁迫; 低温指数 (FD): 温度低于  $8^\circ\text{C}$  的天数, 用于反应极端低温胁迫; 小雨 (R1): 由于小雨对作物产量也有很大影响 (见第二章), 这里定义为降雨小雨 1mm 的日数; 大雨 (R30): 大于 30mm 降雨的日数。

目前作物模型普遍没有充分考虑病虫害 (CPD) 风险, 然而目前大多数模型都没有考虑病虫害模型。Wang 等 (2021a) 基于具有 5,500 多条统计记录的历史数据集, 建立了病虫害风险模型。他们发现中国每个省份的 CPD 发生率都在增加, 全国平均 CPD 发生率增加了四倍。在本世纪末, 气候变化将导致在低排放情景 (SSP126) 下 CPD 发生率增加  $243\% \pm 110\%$ , 在高排放情景 (SSP585) 下增加  $460\% \pm 213\%$ , 其幅度在很大程度上取决于夜间温度升高, 霜冻天数减少。本研究将他们构建的病虫害数据集作为一个训练集, 重新估算了一套病虫害风险, 作为预报因子。使本研究的预测模型可以在一定程度上考虑病虫害风险对产量的影响。通过机器学习分别与不同作物模型建立混合模型可以极大的降低模型的不确定性, 对不确定性起到了约束的作用。

本研究利用机器学习模型 (见第二、三章) 来建立混合模型, 基于观测数据优化了模型参数 (图 5-2), 具体方法见第二章。通过对比不同参数的结果, 玉米的  $n_{tree}$  和  $m_{try}$  为 1100 和 3; 大豆的  $n_{tree}$  和  $m_{try}$  为 900 和 5。



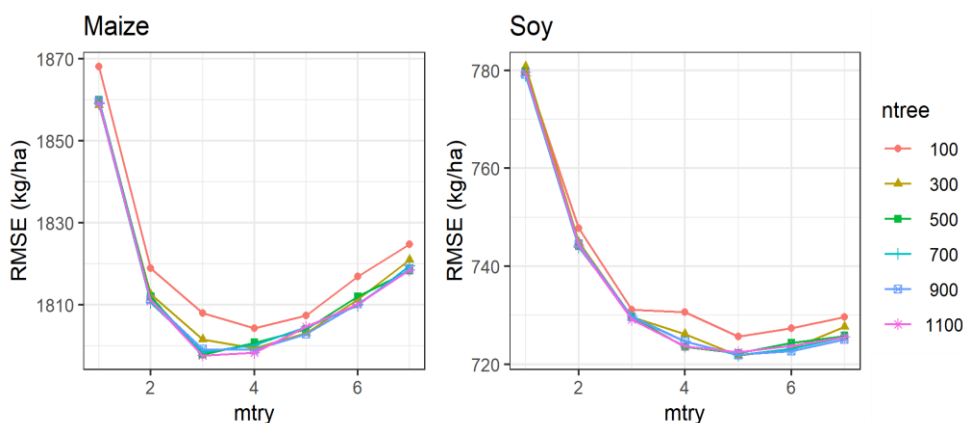


图 5-2 基于玉米和大豆训练数据计算的随机森林模型的不同参数 (mtry 和 ntree) 的 RMSE  
 Fig. 5-2 The cross-validation error rates (RMSE) in tuning parameters (mtry and ntree) for RF using training data sets for predicting maize and soybean yield

由图 5-2 可知, 本研究最终使用的参数。玉米: mtry=3, ntree=1100; 大豆: mtry=5, ntree=900。

将 GGCM 输出的产量数据结合其它预报因子(五个极端气候指数和病虫害风险) 输入机器学习模型, 进行建模并对模型进行验证。然后把未来数据输入该混合模型中进行预测, 未来数据主要包含两个排放情景即 SSP126 和 SSP585。最后对比并分析他们的不确定性来源。本研究用最高和最低的排放情景来更好的反应未来的范围, 具体技术路线见下图 (图 5-3):

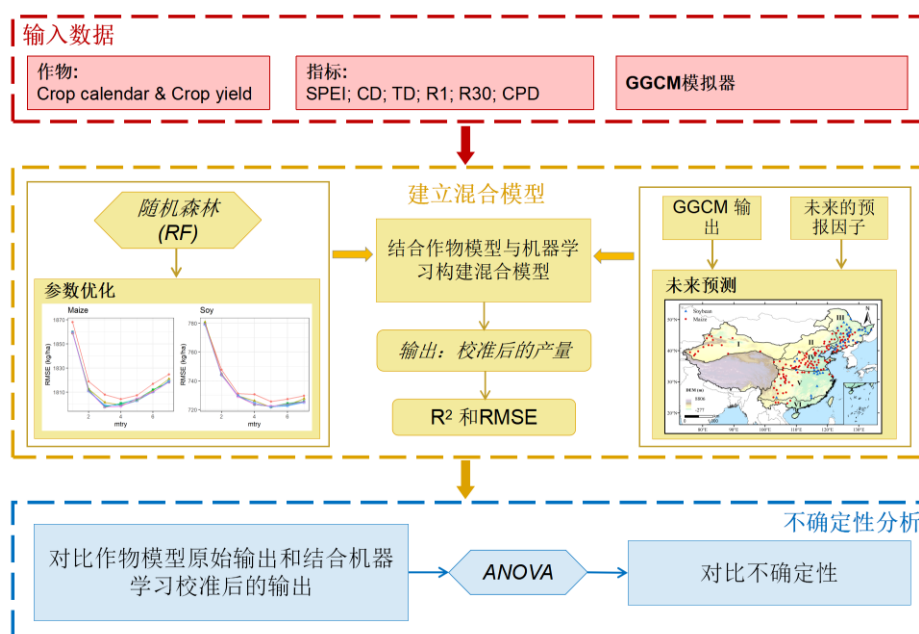


图 5-3 本章构建的不同作物产量预报模型示意图。

Fig. 5-3 Schematic overview of data input and output for the RF model developed in this study.

## 5.2 结果分析

### 5.2.1 历史期间模拟精度

通过交叉验证方法对比了历史期间模型模拟的精度和经过机器学习校准以后的精度（图 5-4）。虽然有 GGCM 应用站点时，可能会极大的降低模型的精度，即使他们的响应与田间实验基本一致（Zhao et al. 2016a）。本研究的结果发现对于 GGCM 模拟的精度较低，其中玉米的  $R^2$  为 0.07-0.37，nRMSE 大约为 0.2-0.4（除了 JULES）；大豆的  $R^2$  为 0.13-0.48，nRMSE 大约为 0.18-0.35。整体上模拟的精度比较低。

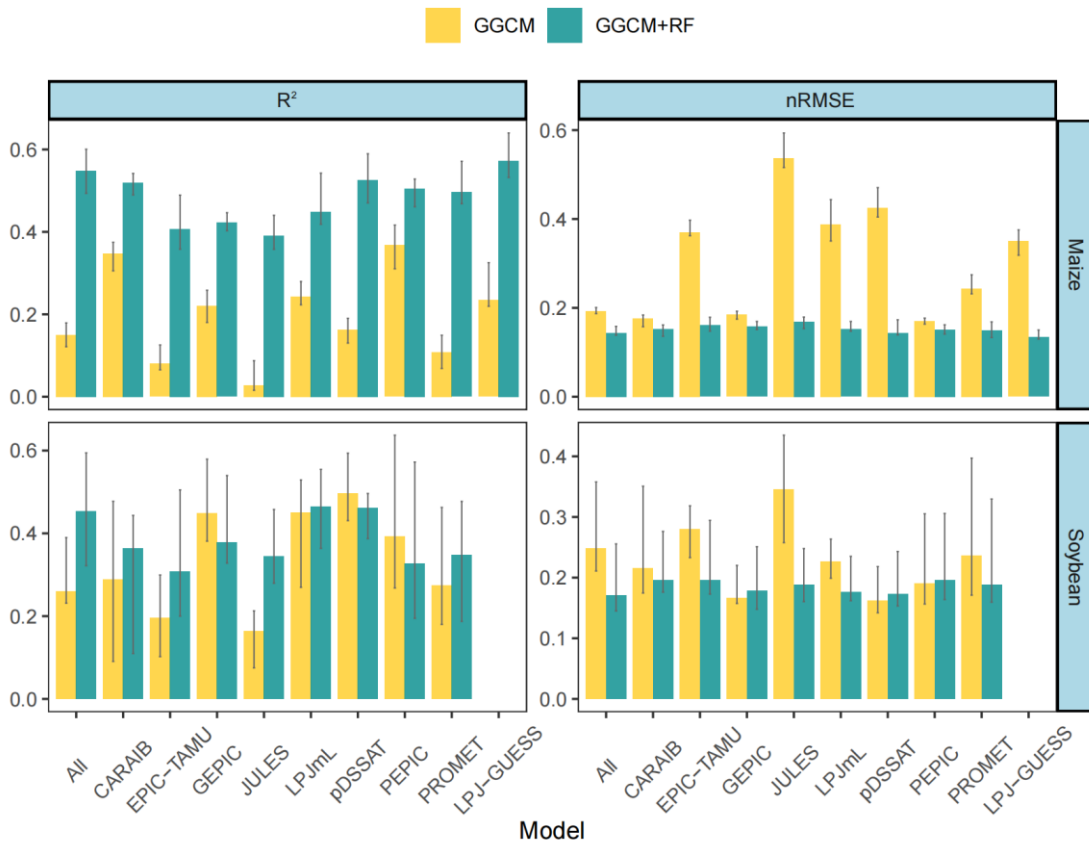


图 5-4 玉米和大豆历史期间（1999-2010）模拟精度评价

Fig. 5-4 Model performance of maize and soybean yield simulation during 1999-2010

虽然 GGCM 对于温度响应基本上可以捕捉到，但是他们对于极端降水的响应不敏感，容易高估极端降水的影响，尤其是极端湿润的情况经常高估产量（Li et al. 2019c）。本研究考虑了不同种类的极端气候，以及病虫害风险。经过机器学习校准后（GGCM\_RF），发现模型精度有了极大的提高，其中玉米的  $R^2$  为 0.37-0.58，nRMSE 全部小于 0.2；大豆的  $R^2$  为 0.31-0.47，nRMSE 在 0.2 以下（图 5-4）。

值得注意的是，可以看到这样的模拟在历史期间精度相比于利用多源环境变量精度要低（Cao et al. 2021a; Cao et al. 2021b; Li et al. 2022a）。这主要是因为本研究只用

了生育期内六个变量和作物模型的输出为预报因子，并没有考虑逐个生育期。本研究这么做的主要原因是如果在模型中输入过多的变量会导致作物模型本身的特性在机器学习中被抵消掉，导致不同模型的结果趋于一致，这样会弱化了每个模型的特点。而输入的变量过少，则导致模型仍然有较大的偏差。因此，为了调整这一个权衡，本研究保证变量的重要性之和小于 50%。此外，如第二、三章所述，在多元环境变量中，遥感数据整体上是最重要的变量，然而本研究主要目的是分析未来气候变化对产量影响，无法获取未来的一些相关的植被指数。

对于玉米来说，无论是 GGCM 还是 GGCM+RF, LPJ-GUESS 的  $R^2$  最高, CARAIB、GEPIC 和 PEPIC 的 nRMSE 最小；对于大豆，PEPIC 和多模型平均的模拟效果更好。然而，JULES 对于玉米和大豆的模拟误差较大，这可能是该模型在气候变化响应中比较敏感，对产量容易造成过高或者过低的模拟。上一章也发现 JULES 往往会带来更多的不确定性。本研究发现多模型平均有时候无法得到最优的模型精度。这可能是由于模型间差异较大，导致多模型平均的结果把一些模型特性低效了，因此一些模型模拟的信息没有反应出来。说明在运用多模型评估产量时，有时候直接把模型平均（如算术平均）可能无法得到比较满意的结果。本研究的结果建议在考虑多模型时，要尽量对模型进行一些选择。以达到最优的模拟或者预测。

### 5.2.2 产量与预报因子的响应关系

预报因子与玉米和大豆产量响应关系曲线如图 5-5 所示。对于玉米来说，低温和病虫害风险为主要的预报因子；对于大豆来说病虫害风险和高温为主要的预报因子。

从响应曲线可以看出，低温对玉米和大豆的产量有一定的促进作用，但当持续时间超过约 20 天时，低温对产量又呈现负向关系。CPD 对产量有明显的负面影响。对于玉米来说，SPEI 与产量呈正相关，说明干旱季节对产量有负面影响，而湿润季节则对产量有正面影响。但对于大豆而言，则是先增加后减小。高温对玉米的影响较为负面，而对大豆则是先增加后下降。R1 对玉米产量有正向影响，但当持续时间超过 13 天后，会转变为负向关系。对于大豆来说，R1 先呈负相关后又变成正相关。而 R30 对玉米产量呈负相关，对大豆则是先正相关，但当持续时间超过 6 天后，也会呈现负向关系。

在 SSP585 情景下，CD 的风险相对较低，而在 SSP126 的预测下，CD 的风险仍然较高。而 CPD 在两种情景下都有很大提升，其中在 SSP585 情景下变化最为显著。高温（TD）在 SSP126 情景下变化不大，但在 SSP585 情景下变化很大，特别是在 21 世纪末。未来 R1 的变化不大（在 SSP126 和 SSP585 情景下），但是 R30 在 21 世纪末在 SSP585 情景下有较大提升。分析不同因子与产量之间的响应关系有助于更好地理解极端气候事件和病虫害风险对产量的影响，也可以提高对混合模型预测的认识。

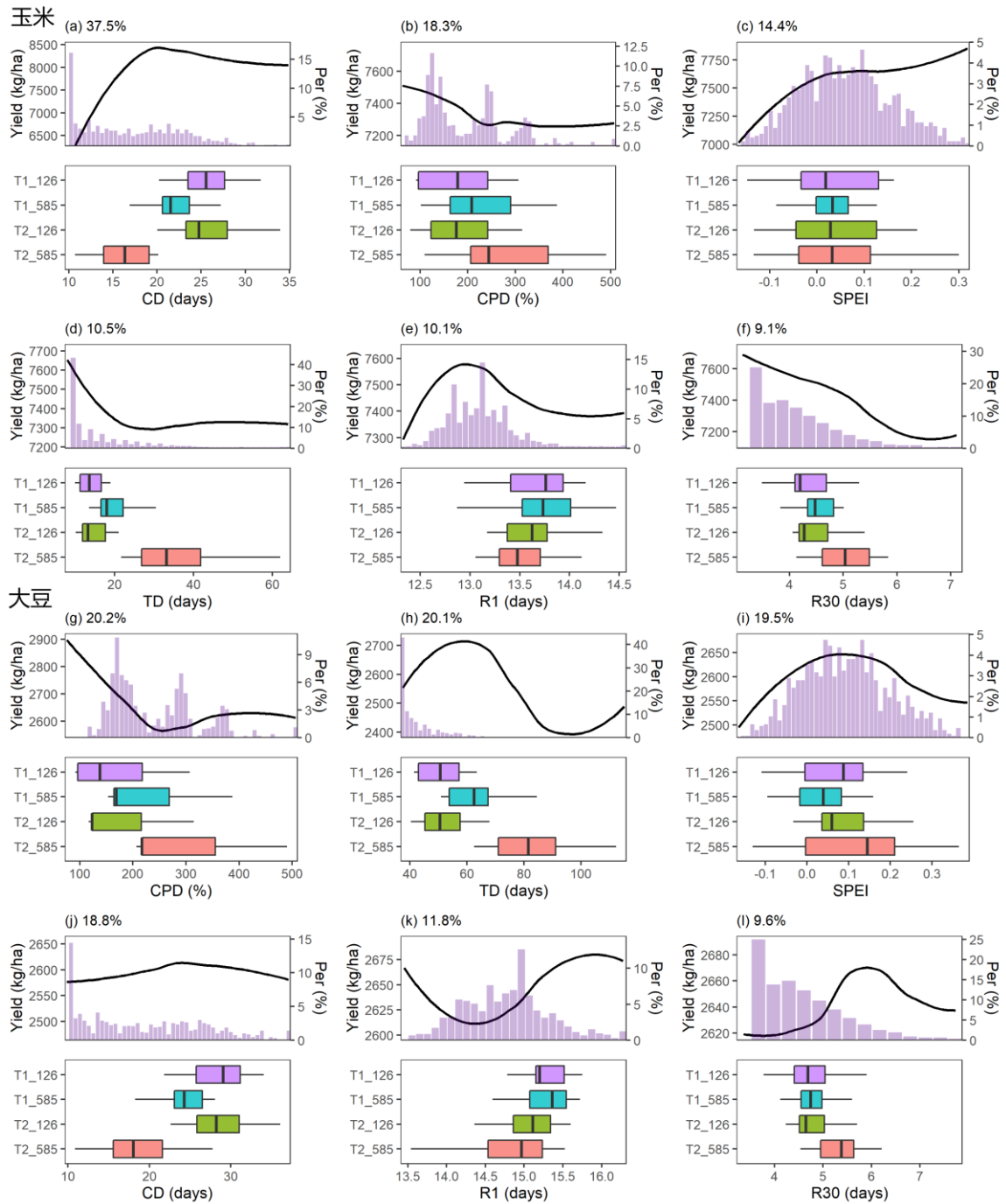


图 5-5 玉米和大豆产量与六个预报因子响应的局部依赖图。图中黑色曲线是经过平滑处理的。直方图展示了每个预报因子的概率分布。箱形图显示了未来基于 21 个 GCM 的两个时段(2040-2069 年和 2070-2099 年)的极端气候事件和 CPD 的发生情况。

Fig. 5-5 Partial dependence of maize and soybean on 6 predictors. The black lines are smoothed representations of the response, with fitted values (model predictions) for the calibration data. The histograms show the probability distribution of the index. The box plots indicate the occurrences of extreme climate events and CPD during two future periods (2040–2069 and 2070–2099) based on the 21 downscaled GCMs.

综合以上分析，可以得出一些结论和建议。对于玉米来说，干旱和低温对产量的负面影响更为明显，此外，高温和 R30 对产量的影响也是负面的。对于大豆来说，干旱和高温对产量的影响更为明显，而 R1 和 R30 则对产量有正面的影响。因此，未来在种植玉米和大豆时，应注意控制生长季节内的干旱和高温事件，以及过多的 R30 天数。另外，本研究还发现，CPD 和 TD 的风险在未来可能会增加，因此，应采取措施降低其对产量的负面影响。

### 5.2.3 气候变化对产量的影响

#### 5.2.3.1 产量预测

本研究分析了多模型集合的产量变化，结果表明，无论是 GGCM 模拟还是混合模型 (GGCM+RF) 的模拟，都显示未来玉米产量存在明显的下降趋势，尤其是在 2040 年之后 (图 5-6a)。对于大豆来说，在 SSP126 情景下，GGCM 预测产量增加，而 GGCM+RF 模型显示自 2040 年左右，产量就开始下降了；在 SSP585 情景下，GGCM 预测的大约 2060 年左右产量开始减产，然而 GGCM+RF 的结果表明，大豆产量减产出现的可能会更早 (图 5-6d)。

虽然两种模拟 (GGCM 或者 GGCM+RF) 都预测了产量下降，但是 GGCM+RF 模拟下降的幅度与 GGCM 有所差别，尤其是对于大豆来说。在图 5-6b 中，发现 GGCM 与 GGCM+RF 模拟的玉米产量变化比较一致，说明 GGCM 对玉米减产模拟的比较好，但是通过混合模型的方法仍然可以对不确定性起到一定的约束作用。此外，作物模型只是根据当下的情况固定了管理措施，而在实际中，可能会采取一些适应措施来减缓气候变化的影响，这就导致了一些极端气候对产量影响的响应曲线可能会变得平缓 (图 5-5)。这些信息可能在机器学习模型中会有体现，而作物模型往往无法反映这些信息。进而导致高估了作物的减产。此外，由于我国多数地区是灌溉农业，灌溉农业可能会减缓高温对产量的影响。因为灌溉时可以降低表层温度，因此，模型可能高估了灌溉区域农业造成的高温胁迫减产 (Siebert et al. 2017; Wang et al. 2021c)。

对于大豆来说，本研究发现 GGCM 对产量可能存在一定的高估。这可能是由于作物模型对于大豆极端事件的考虑不够全面，而机器学习则更能够捕捉这种影响。此外由于大豆与玉米对不同气候因子的敏感性也有差别。例如，相比于玉米来说，大豆产量可能对高温更加敏感，然而一些过程模型对这部分过程过于简化，进而导致了模型可能会低估气候变化对大豆的减产。因此，使用混合模型来分析大豆的产量响应可以更好地预测未来产量的变化趋势，并更好地理解不同因素对产量的影响。总的来说，我们认为基于机器学习构建的混合模型可以更好地考虑到气候变化和作物模型中存在的 uncertainty，从而更准确地预测未来作物产量的变化趋势。

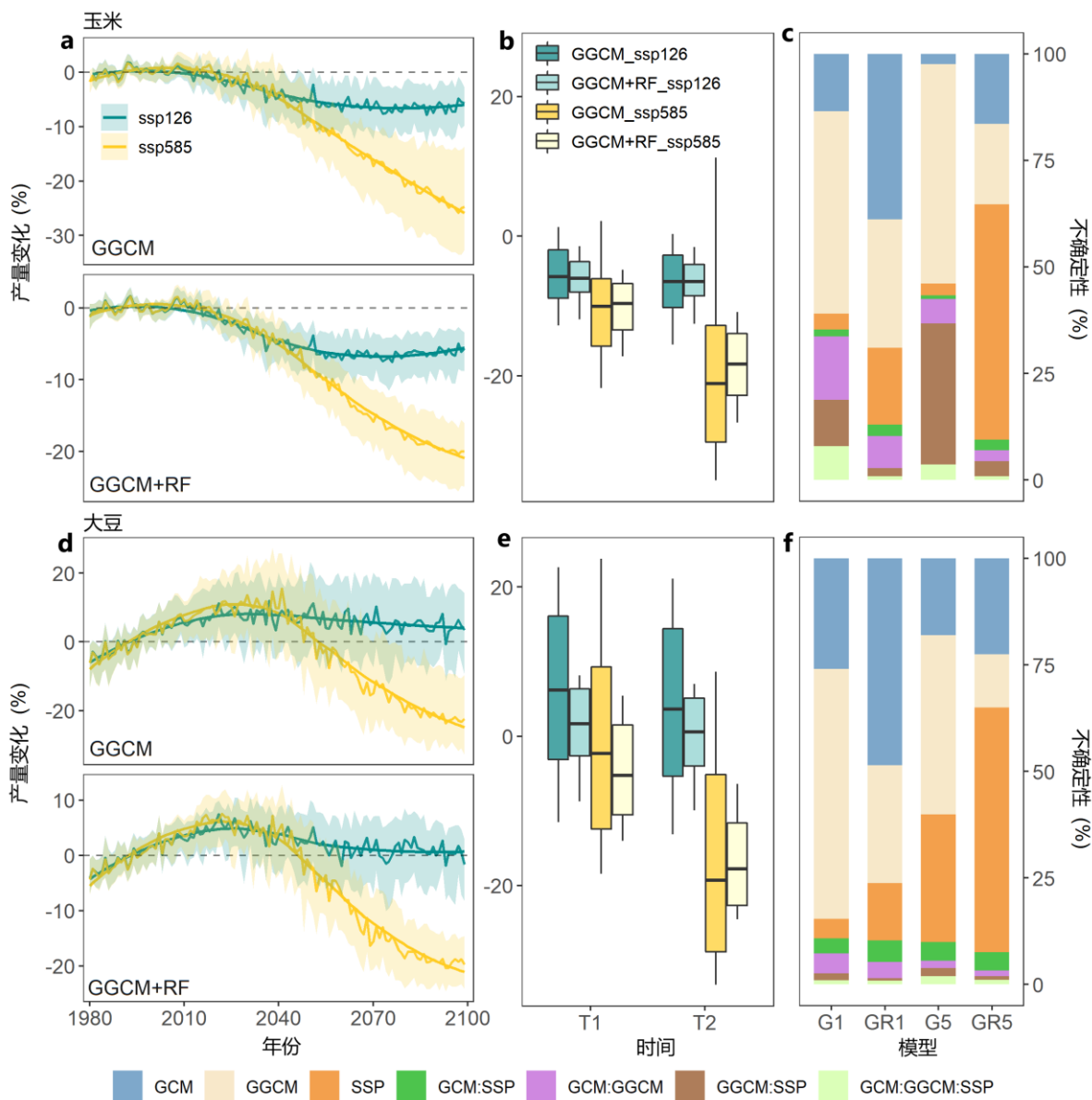


图 5-6 玉米和大豆 1980-2099 期间产量变化的时间序列 (左)。阴影区域反应了 25-75 分位数的范围。2040-2069 和 2070-2099 两个时段产量变化的箱型图 (中)。柱状图产量变化的不确定性来源 (右)。

Fig. 5-6 Time series for maize and soybean yield change during 1980-2099 (left). Shaded ranges illustrate the 25th and 75th percentiles. The box plot shows the maize and soybean yield change during 2040–2069 and 2070–2099 (mid). The bar plot shows the source of uncertainty in maize and soybean yield change projections.

不同区域的产量变化差别较大。玉米在区域 I 和 IV 的产量下降最为明显。整体上 GGCM+RF 模拟的下降幅度略小于 GGCM，尤其是在区域 V。对于大豆，区域 IV 和 V 下降最为明显。在 SSP126 情景下，GGCM+RF 模拟的下降幅度略大于 GGCM。



在 SSP585 情景下，区域 IV 和 V 的 GGCM 模拟的大豆产量变化下降更加明显。在区域 III 中，在 SSP126 和 SSP585 情景下，GGCM+RF 模拟的产量下降都更为明显。具体的历史和未来产量变化的结果以及区域差异见图 5-7 和 5-8。

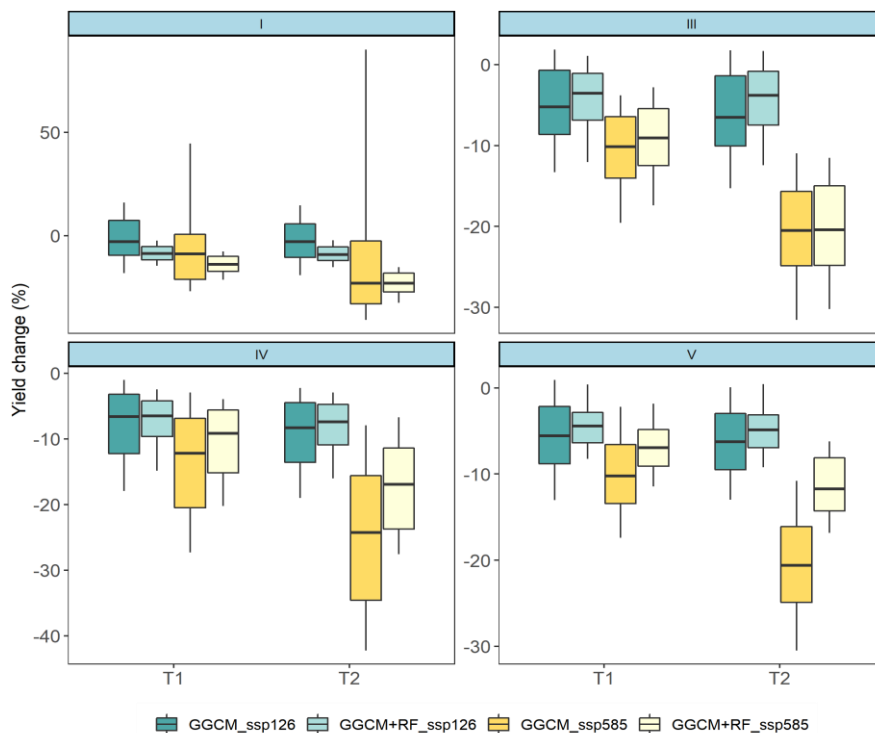


图 5-7 玉米 2040-2069 和 2070-2099 两个时段不同区域 GGCM 和 GGCM+RF 产量变化的箱型图  
 Fig. 5-7 The box plot shows the maize yield change of GGCM and GGCM+RF during 2040–2069 and 2070–2099 in different sub-regions

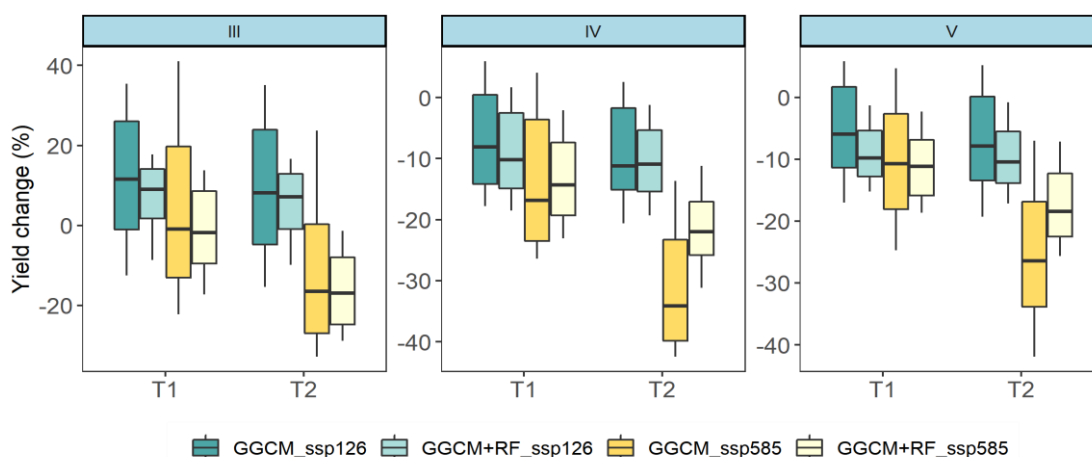


图 5-8 大豆 2040-2069 和 2070-2099 两个时段不同区域 GGCM 和 GGCM+RF 产量变化的箱型图  
 Fig. 5-8 The box plot shows the soybean yield change of GGCM and GGCM+RF during 2040–2069 and 2070–2099 in different sub-regions

### 5.2.3.2 不确定性分析

本研究分析了 T1 和 T2 两个时段玉米和大豆产量不确定性的来源。在 T1 时段，不论校准前还是校准后，玉米产量的主要不确定性来源是作物模型，而在 GGCM+RF 的模拟中，气候模式所占比例下降，作物模型所占比例增加；而在 T2 时段，排放情景成为主要不确定性来源（图 5-6）。大豆在 T1 时段的不确定性来源是作物模型和气候模式，T2 时段则是排放情景。矫正以后，排放情景在大豆的不确定性来源中占比更大（图 5-6）。箱型图显示，与 GGCM 模拟相比，GGCM+RF 的范围明显缩小，说明基于机器学习校准可以在一定程度上约束不确定性。对比约束前后不确定性来源的差异，在全国尺度上，GGCM+RF 模拟的可以降低作物模型造成的不确定性的比重。

不同区域的不确定性来源存在很大差异，说明不确定性来源因地制宜（图 5-9 和图 5-10）。对于玉米来说，在区域 I 和区域 IV 作物模型是主要的不确定性来源，而在区域 III 和 V，在 T1 时段，气候模式占的不确定性来源的比重更大，而 T2 阶段排放情景占的比重更大。对于大豆来说，区域 III 中作物模型是主要的不确定性来源，而在区域 IV 和 V，气候模式或者排放情景为主要的不确定性来源。玉米不同区域约束前后的主要不确定性来源变化不大。然而，对于大豆来说却有明显变化，尤其是 T2 时段。分析不同不同区域约束前后不确定性来源的差异，可以更好的根据不同区域，提出降低不确定性的策略，可以为精细化农业管理和调整农业政策提供参考依据。

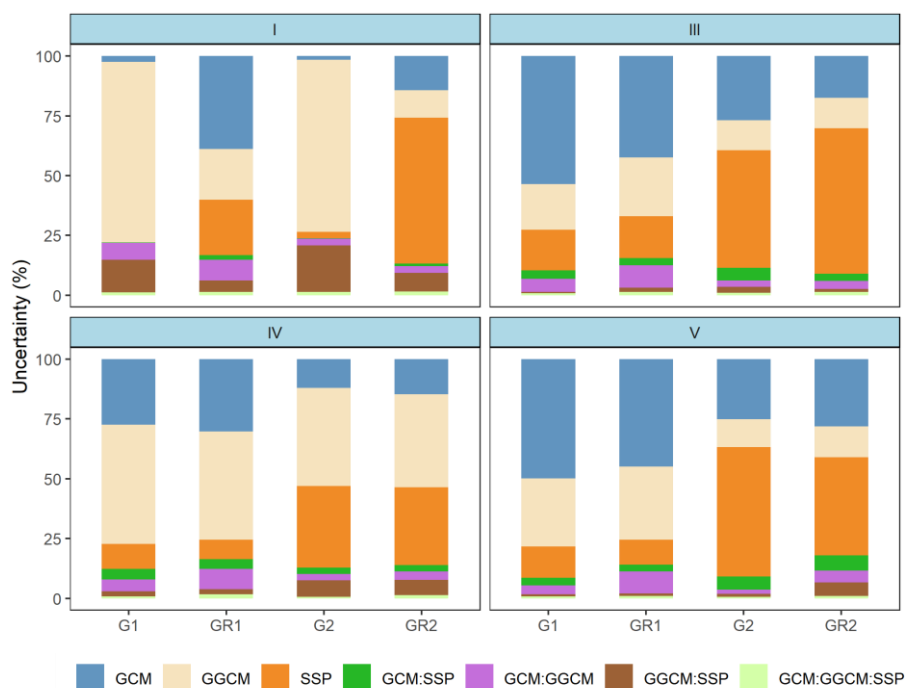


图 5-9 玉米不同区域未来产量变化的不确定性来源。

Fig. 5-9 The bar plot shows the source of uncertainty in maize yield change projection in different sub-regions.

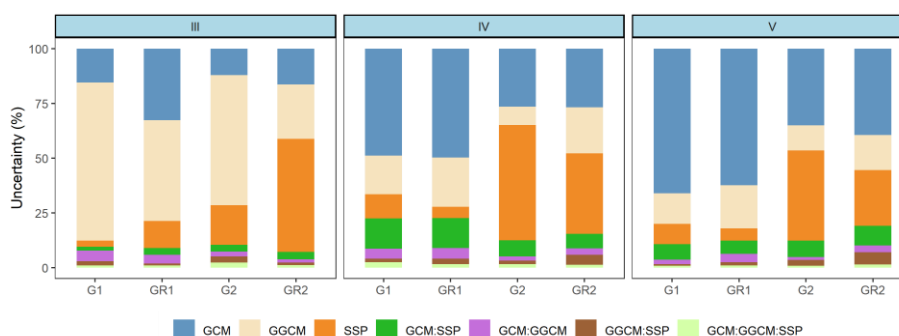


图 5-10 大豆不同区域未来产量变化的不确定性来源。

Fig. 5-10 The bar plot shows the source of uncertainty in soybean yield change projection in different sub-regions.

## 5.3 讨论

### 5.3.1 预报因子与产量的响应关系

本研究揭示了极端气候与病虫害风险、产量之间的关系（图 5-5）。病虫害风险与产量呈现明显的负相关，但无论是玉米还是大豆，在风险大于约 220 的情况下，产量不再下降。这可能是由于该数据集存在一些异常值，同时，高病虫害风险下的数据也存在一些未知的影响因素。对于玉米和大豆，它们的低温响应都表现为先增大后减小的趋势，而这种响应与小麦相似（见第二章）。当低温超过一定阈值时，可能会对玉米和大豆的产量造成一定的减损。这可能是由于春夏季的低温往往伴随着太阳辐射的下降，从而导致作物减产（Gao et al. 2022）。高温和干旱是最常见的导致产量减产的因素（Guo et al. 2019; Li et al. 2019c），且玉米和大豆都对高温比较敏感，这可能是由于谷物败育、籽粒大小潜力下降和籽粒生长减慢的混合作用（Hoffman et al. 2020）。

对于玉米而言，随着 SPEI 指数的增加，玉米产量也随之增加，说明水分对玉米的产量具有较大的决定性作用。然而对于大豆而言，当 SPEI 指数较大时，产量仍然下降，这可能是由于极端潮湿也会对产量造成减损，而这部分减损常常被作物模型所忽略（Li et al. 2019d）。这也可能是本研究中作物模型总体高估大豆产量的原因之一。此外，过多的小雨也被证明对产量具有较大的负面影响（Lesk et al. 2020）。本研究的结果也表明，当小雨天数超过 13 天时，产量会下降。这可能是由于过长时间的小雨会带来长期的低温和辐射压力。暴雨对玉米的影响比大豆更明显，这可能是由于大豆在苗期对涝渍比较敏感，而暴雨在七至八月份更为常见。而暴雨的往往伴随着大风，在暴雨等强对流天气中，玉米更容易发生倒伏，或病虫害等。

通过基于机器学习的分析，本研究可以深入了解产量形成的机理和识别关键因素，从而制定更有效的农作物管理措施来提高产量和品质。机器学习方法能够自动挖掘大

量数据中的复杂非线性关系，识别对产量有重要影响的预报因子，并预测它们对产量的贡献。这些预测结果可以帮助优化农作物的管理措施，提高产量和品质。

### 5.3.2 气候变化导致玉米和大豆减产

与过去的研究结果类似，气候变暖（例如升温）将导致玉米和大豆减产（Huang et al. 2022; Kothari et al. 2022）。尤其是当 21 世纪中叶以后，产量将大幅度下降。因此，如果不采取适应措施，气候变化预计会降低农业生产力并增加作物歉收的风险。要在 2040 前改进新的品种以适应生长季（图 5-11）。假设的适应性品种对稳定产量有很大的作用如图 5-11 产量在 21 世纪末才开始下降，在 SSP585 情景中预计超过一半地区需要改进适应性品种（Zabel et al. 2021）。除了品种适应之外，其他适应措施，例如改变播种日期（Huang et al. 2020; McDonald et al. 2022），在目前生长期受低温限制的地区利用更长生长期的潜力，或转向不同的作物，都可能是可行的适应措施，且未来粮食种植带有向两级地转移的趋势（Franke et al. 2022）。虽然气候变化对我国玉米和大豆的粮食生产有很大的不利影响，但是目前我国粮食生产仍有较大的潜力，通过优化管理措施可以在提高玉米和大豆产量的同时减少施肥量以及温室气体排放，实现可持续的增产（Liu et al. 2021b）。总的来说，本研究的结果可以帮助农业从业者和政策制定者更好地了解气候变化对农业的影响，以及如何适应这些变化。例如改善土地利用和管理、提高农作物的抗病虫能力、减少化肥和农药使用等措施，以维护农业的可持续性。

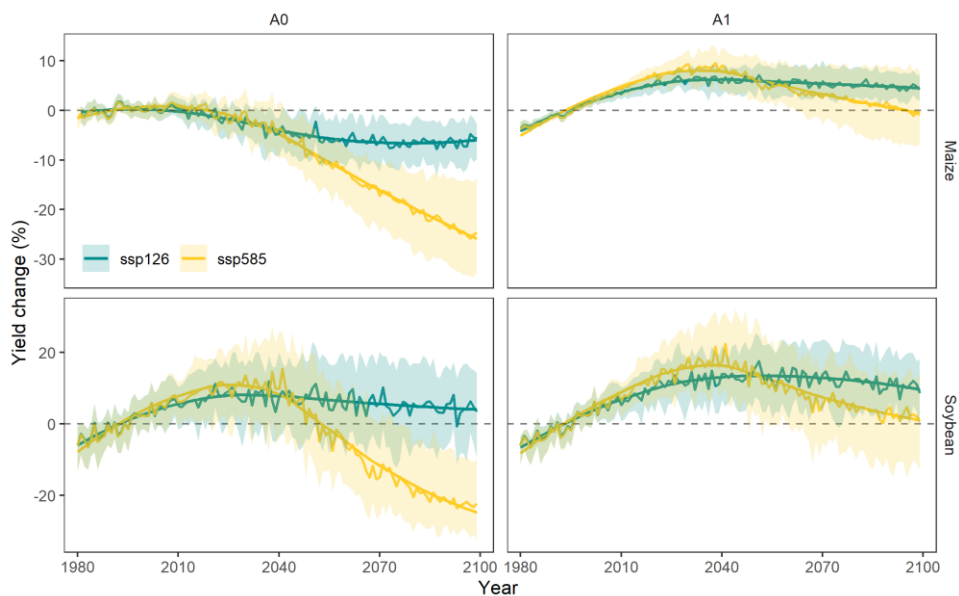


图 5-11 考虑品种的（A1）玉米和大豆 1980-2099 期间产量变化的时间序列变化。

Fig. 5-11 Time series for maize and soybean yield change considering the variety (A1) during 1980-2099.

### 5.3.3 机器学习约束不确定性

由于当前作物模型存在较大的不确定性，因此提高预测的稳健性成为一个重要的科学问题。过去，一些研究基于观测数据来约束模型的不确定性。例如，Zhao 等(2016b)比较了水稻产量对田间变暖实验和三种建模方法（统计模型、本地作物模型和全球网格作物模型）的温度升高的敏感性。研究发现，相比于实验和本地化的作物模型，统计和全球网格作物模型都表明变暖对产量的负面影响较小。为了约束大范围的全球网格化作物模型结果，研究人员使用了来自田间变暖实验的数据，提出了一种条件概率方法。Wang 等(2020e)将小麦、玉米、水稻和大豆的全球田间变暖实验数据集（48个站点）与网格化全球作物模型相结合，以生成关于作物产量对温度的变化影响的田间数据约束估计值紧急约束方法。研究发现，模型不确定性可以降低 12-54%。然而，不同的响应过程都会带来不确定性。本研究通过机器学习考虑了不同的响应，可以更高效地约束产量预测的不确定性。研究结果表明，通过结合多源数据与机器学习，产量预测的不确定性可以降低 25.8-74.1%。通过比较机器学习约束前后的不确定性变化，可以帮助评估机器学习在农业预测中的应用价值。约束后的预测结果更加一致，那么可以进一步探究如何将机器学习方法与传统的作物模型相结合，以提高预测结果的准确性。例如将机器学习捕捉到的气候与产量的非线性关系嵌入作物模型，来改进模型使其更好的捕捉因子与产量的响应关系。

### 5.3.4 不确定性来源

分析不确定性来源对于制定减少不确定性策略方面非常重要。本研究结果显示，对于玉米来说，作物模型是不确定性的主要来源，而对于大豆来说，气候模式则是主要因素。在 T2 时段，排放情景成为了主要的不确定性来源，可能是因为本研究选择的两个排放情景（SSP126 和 SSP585）是最高和最低的情景，它们贡献了较大的方差。但是，本研究没有考虑到适应措施，例如品种和播期的变化。Huang 等(2022)研究了中国玉米气候变化适应性的主要不确定性来源，结果表明当考虑不同播种日期时，气候模型是主要的不确定性来源，而当考虑改变品种时，作物模型则是主要的不确定性来源。

不同区域存在巨大差异的不确定性主导因素。本研究发现，在四个玉米种植区域和三个大豆种植区域中，不确定性的主导因素存在显著的差异。例如，在第三个大豆种植区，作物模型是产量变化不确定性的主导因素，而在区域 V 则是气候模式。这表明不确定性的来源在不同地区是多种多样的，并可能由当地的气候、土地利用、作物品种和灾害风险等管理因素决定。这些因素的不同会对作物产量产生不同的影响，因此导致不同区域的不确定性来源存在显著差异。因此，这些因素在空间上也存在着明显的分布规律，如何揭示这些空间差异是一个重要的科学问题。此外，本研究发现，

多模型平均结果并不总是最佳的选择,因此在选择模型时,应该充分考虑模型的特点,选择最适合分析问题的模型。未来的工作将集中于探索不确定性来源的空间分布以及不同模型的组合。

### 5.3.5 研究的局限性

尽管本研究采用了一定的方法来约束多模型预测的不确定性,但是仍存在一些局限性。首先,本研究主要采用在外部耦合机器学习的方法来提高模型预测的精度,然后约束不确定性。然而,本研究没有针对作物模型的结构进行改进,例如温度响应(Wang et al. 2017b),光合(Wang et al. 2022b)或涝渍(Liu et al. 2023)影响等方面的改进。其次,机器学习是数据驱动模型,数据质量以及数量对模型精度有非常大的影响。由于数据收集的限制,本研究所使用的观测数据比较少,且数据质量可能存在一定程度的偏差,这可能会对模型预测结果产生影响。此外,由于病虫害等因素的数据比较糙,可能会对结果有较大的影响。Wang等(2021a)构建了主要粮食作物的病虫害风险(包括玉米、小麦和水稻),并没有大豆。本研究基于他们的数据集构建了一个比较粗的训练集,重新模拟了大豆的病虫害风险,这里的数据具有较大的偏差。然而,目前确实没有更加详细的病虫害数据集。这里值得注意的是,存在一些病虫害风险的值大于100%,这是因为这套数据集是根据月尺度建立的,所以可能会超过100%。例如,一个100公顷的土地,三月份发生的面积为60公顷,四月份发生的面积为40公顷,五月份发生的面积为20公顷。那么这个区域的病虫害风险则为120%。第三,因为不同的作物模型对气候变化响应过程差异很大,因此,不同模型的组合和集合数量对不确定性来源也有很大的影响。例如,集合中选择三个相似的作物模型,和三个差异较大的气候模式,那么得到的不确定性来源可能是气候模式大于作物模型,反之亦然。在上一章中探讨了不同组合对于不确定性的影响,然而本研究没有对比对模型进行约束后,模型数量以及组合对产量预测以及不确定性影响,分析这些因素可以进一步提高对气候变化影响的认识。最后,由于网格模型在田间模拟仍然存在较大的不足,例如原始精度较差(即使经过站点校准)。虽然通过机器学习可以实现较好的模拟,但是精度仍然不高。本研究虽然提出了一个全新的框架来约束并探讨不确定性。但是将本章工作通过过程模型在田间尺度进行参数优化以及校准后再进行构建混合模型可能会效果更好,后续将开展类似的工作以填补这些不足。

## 5.4 小结

本研究以玉米和大豆为例结合网格化作物模型以及机器学习考虑不同因素(极端气候与病虫害风险),构建混合模型并分析了历史和未来不同排放情景下气候变化对玉米和大豆产量的影响。然后分析了它们的不确定性来源。主要结论如下:



(1) 通过结合机器学习来校准作物模型输出, 可以极大的提高模型的精度。玉米的  $R^2$  为 0.06-0.37, nRMSE 大约为 0.2-0.4(只有 JULES 大于 50%, 说明误差较大); 大豆的  $R^2$  为 0.12-0.48, nRMSE 大约为 0.19-0.35。校准后玉米的  $R^2$  为 0.37-0.58, nRMSE 全部小于 0.2; 大豆的  $R^2$  为 0.31-0.47, nRMSE 在 0.2 以下。

(2) 病虫害、高温和干旱为大豆产量的主导因子; 低温、病虫害和干旱为玉米的产量变化的主导因素。

(3) 本研究的方法可以约束 25.8-74.1%的不确定性。对于 GGCM, 作物模型是主要的不确定性来源, 对于 GGCM+RF, 在 T1 阶段 GCM 是主要的不确定性来源, 而 T2 阶段, SSP 为主要的不确定性来源。

本研究改进了网格模型的预测, 并有效的约束了作物模型的不确定性。同时揭示了不确定性来源的区域差异。可以为当地农民和政策制定者提供一些有用的信息, 帮助他们制定应对气候变化下极端气候增加的适应战略。

## 第六章 未来情景下气候驱动因子对作物产量的影响

尽管作物模型预测仍然存在一些不确定性 (Kamali et al. 2022), 但是已经被广泛应用于分析气候变化对农业的影响等研究 (Hasegawa et al. 2021; Heino et al. 2023; Minoli et al. 2022)。由于区域的气候情况受大气环流场的影响, 因此, 直接分析气候驱动因子对产量的影响, 有助于帮助我们更好地理解气候变化对产量影响的机制以及对未来粮食产量趋势的预测。粮食生产和粮食安全已成为全球公共健康问题, 并且随着人口和粮食需求的不断增加而更加重要 (Lam et al. 2013; Wheeler and Braun 2013)。然而, 全球变暖 (通常导致极端干旱、暴雨和热浪等灾害) 可能会影响粮食供应, 并威胁整个粮食系统的稳定性 (Gohar and Cashman 2016; Ray et al. 2015)。最近的研究表明, 这些风险可能会在多个粮食产区中同时发生, 尤其是在最近十年更加频繁 (Gaupp et al. 2019)。

当前, 大气环流因素已经成为影响作物产量的一个关键因素 (Anderson et al. 2023; Cao et al. 2023; Jiang and Zhou 2023), 大约三分之二的粮食产区受到厄尔尼诺南方涛动 (ENSO)、印度洋偶极子 (IOD)、北大西洋涛动 (NAO) 等大尺度的大气环流场的影响 (Heino et al. 2018)。这种影响在未来气候变化下可能会进一步加剧。因此, 迫切需要更好地量化未来气候变化下作物产量与气候驱动因子的响应关系。本章将统计模型结合作物模型的方法进一步扩展, 以更好地理解气候驱动因子对作物产量的影响。主要研究的目的是: (1) 识别和比较历史和未来全球作物产量变化的主要气候驱动因素; (2) 可靠地估计全球和区域作物产量对气候驱动因素的敏感性, 并量化历史和未来大气环流因子强振荡阶段对产量的影响; (3) 分析该研究的不确定性以及主要不确定性来源。本研究可以提高对气候变化冲击全球主要粮食产区的理解, 这对提高全球粮食系统的韧性具有潜在的影响。

### 6.1 材料和方法

#### 6.1.1 大气环流指数

气候驱动因子波动对不同区域产量的影响不同 (Heino et al. 2018)。因此, 本研究基于 Tian 等 (2019) 的分区, 将全球划分为了 18 个气候分区 (图 6-1)。其中包括: 包括巴西 (BRA)、中美洲 (CAM)、加拿大 (CAN)、中亚 (CAS)、中国 (CHN)、赤道非洲 (EQAF)、欧洲 (EU)、韩国和日本 (KAJ)、中东 (MIDE)、北非 (NAF)、南美洲北部 (NSA)、大洋洲 (OCE)、俄罗斯 (RUS)、南非 (SAF)、南亚 (SAS)、东南亚 (SEAS)、南美西南部 (SSA), 美国 (USA)。由于数据比较稀疏, 本研究没

有考虑 NAF 和 MIDE 区域。

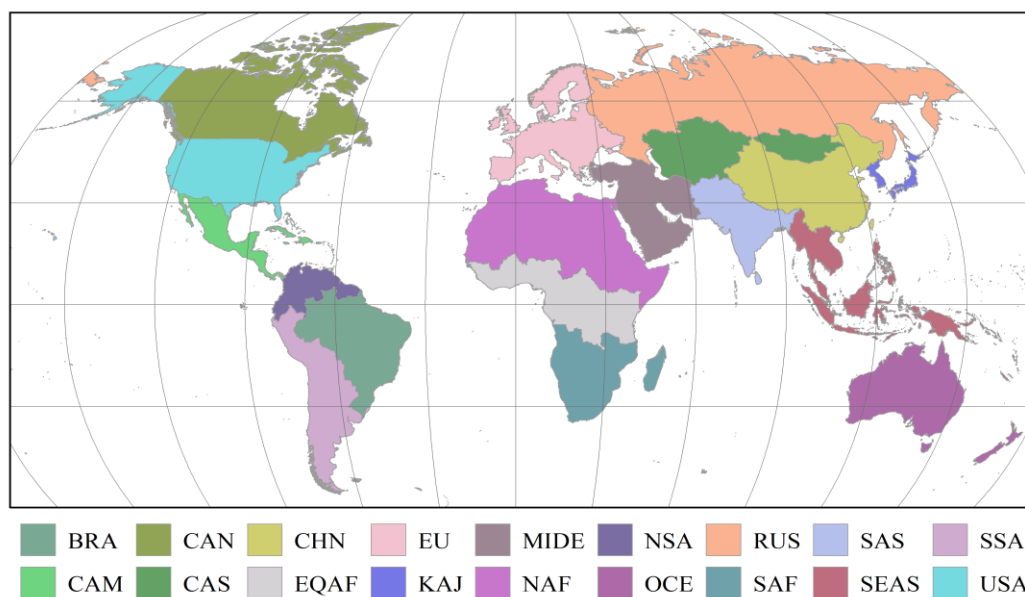


图 6-1 本研究所用的全球 18 个分区空间分布

Fig. 6-1 Spatial patterns of the 18 sub-regions in this study

本研究使用了三个主要的大气环流因子：IOD、ENSO 和 NAO。本研究基于五个 GCM 预测未来这些因子的变化。Nino3.4 指数是通过 Niño3.4 区域（170°W-120°W，5°S-5°N）内的平均 SST 异常计算的（Rayner 2003）。ENSO 是驱动全球变化的主要因子之一（McPhaden et al. 2020）。虽然 ENSO 是在热带太平洋形成的，但其影响范围是全球性的，包括社会经济（Anttila-Hughes et al. 2021; Hsiang et al. 2011）、生态系统（Le et al. 2021）和粮食生产（Iizumi et al. 2014）。鉴于 ENSO 已显示出强烈影响全球作物产量，在气候变化下，ENSO 对作物产量的影响可能也会有所变化。

IOD 由赤道西印度洋（50°E-70°E，10°S-10°N）和赤道东南印度洋（90°E-110°E，20°S-0°N）之间的异常 SST 梯度表示（Saji and Yamagata 2003）。IOD 也是影响极端降水和干旱风险的主要气候驱动因素（Saji et al. 1999）。强 IOD 阶段独立影响区域气候情况（可能会导致洪涝、干热等事件），进一步影响作物光合作用（Wang et al. 2021b）和作物产量（Feng et al. 2022b）。例如，在 2019-2020 年期间，pIOD 引发的干热天气在很大程度上导致了澳大利亚的森林火灾（Phillips and Nogrady 2020）。

根据之前的研究（McKenna and Maycock 2021; Stephenson et al. 2006），本研究将 NAO 定义为北大西洋南部地区（90°W-60°E，20°N-55°N）和北部地区（90±60°E，55°N-90°N）之间的面积平均 SLP 差异。与基于统计的指数相比，该方法对观测值和模型之间行动中心的差异不太敏感。这项研究包括 NAO，因为它特别影响北半球的气候，如欧洲和北美，这通常与极端事件（炎热或潮湿）、积雪和风有关（Bastos et al.

2016)。因此,NAO 进一步影响作物生长和产量,特别是在北半球(Anderson et al. 2019; Heino et al. 2020)。

有几种气候模式(例如 PDO、AMO 和 AO)可能也会显著影响全球气候模式和极端事件,然后进一步影响全球作物产量(Nouri et al. 2021; Schillerberg et al. 2019)。然而,在本研究中,主要关注年际变化,而十年气候变化不适合本研究。因此,本研究仅考虑 ENSO、IOD 和 NAO。此外,许多先前的研究发现,此类气候指数在很大程度上影响了全球作物产量(Anderson et al. 2019; Anderson et al. 2021; Heino et al. 2018)。本研究主要分析两个时段(历史:1901-1999;未来:2001-2099),并定义当振荡相位高于(或低于)第 75 (25)百分位时为强震荡阶段。

### 6.1.2 产量变化的主要影响因子

本研究使用具有三个气候输入随机森林(RF)模型来识别不同作物产量的主要气候驱动因素。RF 模型是一种基于多元回归树的流行算法,可用于分析不同因素之间的非线性响应。在这项研究中,在 RF 模型中使用了生长季节气候驱动因素(来自五个 GCM)和作物产量(12 个 GGCM 集合)。本研究将重要性值归一化为 100%,然后判断他们的显著性。具有最大重要性值的因素被确定为影响作物产量的主导因素。当标准差(ENSO、NAO 和 IOD 之间的标准差)低于 10%时,确定这些地区具有一个以上的主要气候驱动因素。通过 R 中的“rfPermute”包计算重要性值和显著性水平(Klueter et al. 2015)。

通过 RF 模型计算气候因子与作物产量的部分依赖关系,得到了气候因子与作物产量的响应函数。部分相关图可以揭示气候驱动因素与作物产量之间的关系是线性的、非线性的,还是更复杂的(Friedman 2001)。在这项研究中,使用了“pdp”R 包为每个网格建立 RF 模型,并使用偏依赖函数量化了作物产量在强负(正)期和中性期之间的变化。

### 6.1.3 敏感性分析

虽然随机森林模型可以量化强振荡阶段的作物产量变化,但在全球范围内很难反应作物产量对气候驱动因素的响应关系。为了弥补这一缺陷,本研究采用动态线性模型(DLM)来捕捉作物产量对气候驱动因素的敏感性。相对于机器学习模型,贝叶斯 DLM 是一种统计模型,能够更好地捕捉不同时间间隔的多变量时变关系(Prado and West 2010)。虽然与机器学习模型相比,DLM 降低了预测精度,但它能够以更易于解释的方式提供有用的信息。在此研究中,使用 DLM 对每个点进行建模,考虑作物产量(包括小麦、玉米、水稻和大豆)和不利的生长季节气候驱动因素。该方法可以表征作物生长季节内作物产量对 IOD、ENSO 和 NAO 的敏感性变化。本章使用 R 语言的“dynlm”包进行 DLM 建模(Zeileis et al. 2005)。此外,先前的研究表明,DLM 在

量化气候变化对作物生产的影响方面具有广泛的应用价值 (Hein et al. 2018; Liu et al. 2019; Zhang et al. 2021; Zhang et al. 2022)。

根据 IPCC 第五次评估报告的指导意见, "非常不可能"、"不可能"和"可能"这些术语分别代表了结果可能性的 0-10%、0-33%和 66-100% (Le et al. 2021; Stocker et al. 2013)。综合本研究结果, 本研究重点考虑  $P < 0.33$ , 即强振荡阶段对这些地区的作物产量显示出没有因果关系的可能性较低。因此, 我们认为当  $P < 0.33$  时, 气候波动和作物产量之间存在着因果关系。同样地, Le 等 (2021) 分析了 ENSO 对碳汇的影响时, 也采用了  $P < 0.33$  的显著性水平。Cao 等 (2023) 分析了历史期间 ENSO 对产量的影响时, 使用了更高的显著性水平 ( $P < 0.1$ )。这导致了结果中显著减产的区域比较稀疏, 可能忽略了一些可能对产量影响较大的地区, 例如美国地区的玉米和冬小麦 (Schillerberg et al. 2019; Schillerberg and Tian 2020), 中国的玉米等 (Shuai et al. 2016)。

本章的技术路线如下图所示:

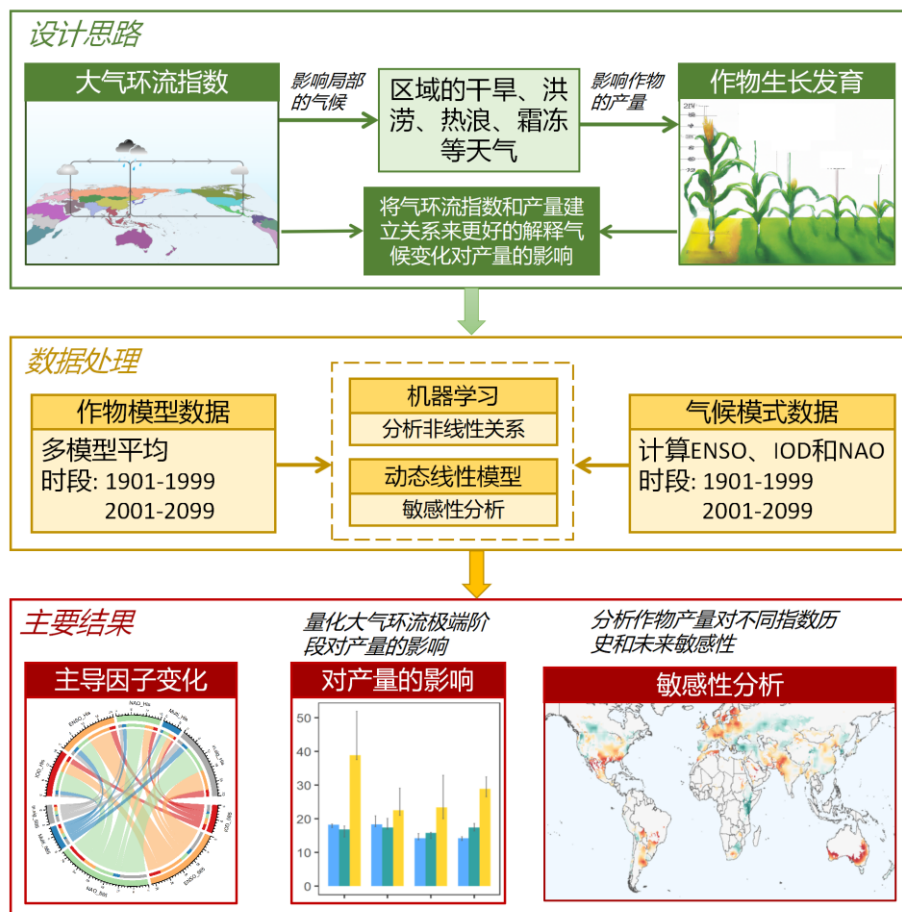


图 6-2 本章未来气候变化下大气环流因子对产量影响分析的技术路线图

Fig. 6-2 Schematic overview of climate drivers impact on crop yield projections under future climate change in this study



## 6.2 结果分析

### 6.2.1 未来 IOD、ENSO 和 NAO 的预估

本章预测了未来 IOD、ENSO 和 NAO 的变化(去趋势),图 6-3 显示了 IOD、ENSO 和 NAO 的方差。ENSO 的方差中,除了 UKESM1-0-LL 下降其余的都是增加的。IOD 的方差变化中, GFDL-ESM4 和 MRI-ESM2-0 的方差下降,其余的上升 (IPSL-CM6A-LR、MPI-ESM1-2-1HR 和 UKESM1-0-LL)。NAO 的方差在 SSP126 排放情景中为上升(只有 IPSL-CM6A-LR 下降),在 SSP585 排放情景中下降。整体上, ENSO 的方差上升, IOD 的变化不大, NAO 有略微的下降。具体见图 6-2。同时分析了 IOD、ENSO 和 NAO 的时间序列变化如图 6-4 所示。图中可以看出 NAO 的模型间的差异小于 IOD 和 ENSO 的。

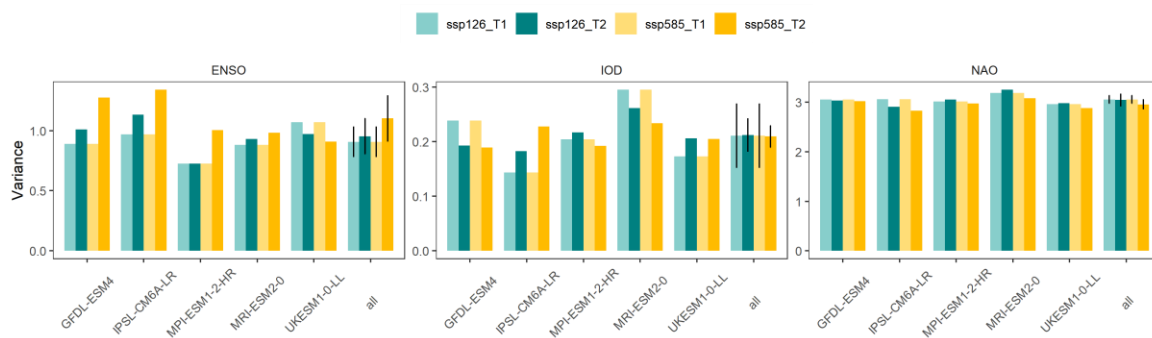


图 6-3 IOD、ENSO 和 NAO 的方差预测

Fig. 6-3 An inter-model consensus on IOD, ENSO and NAO variance change

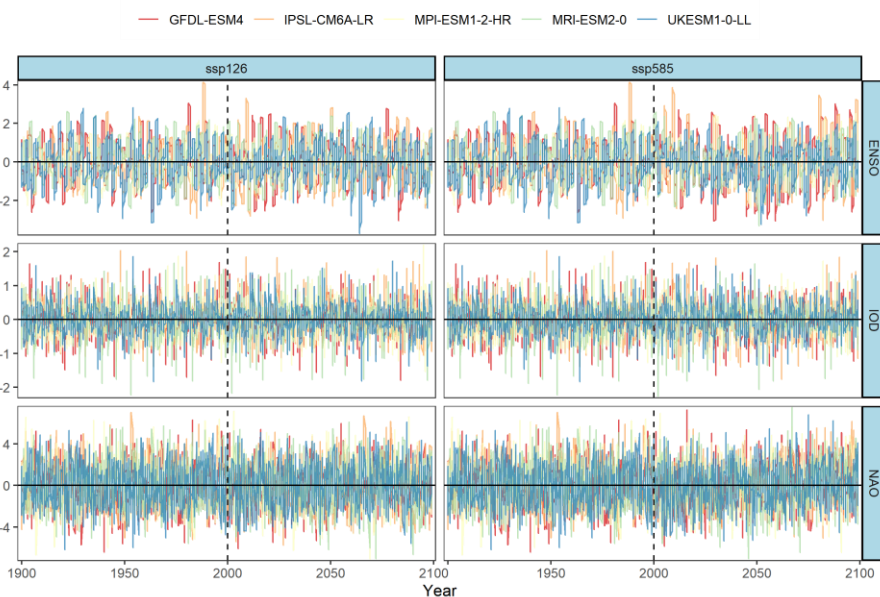


图 6-4 IOD、ENSO 和 NAO 的时间序列变化

Fig. 6-4 An inter-model consensus on IOD, ENSO and NAO time series variation



## 6.2.2 历史和未来产量变化及其与环流指数的显著性关系

本节研究了历史和未来两个时段的作物产量变化（图 6-5a）。结果表明，未来玉米产量有所下降，而小麦和大豆产量将有所增加，而水稻产量变化不大。这可能是由于不同模型之间存在较大差异，导致多模型平均结果相互抵消（例如上一章，图 5-8）。值得注意的是，未来各作物的变异系数都将增加（图 6-5b），这表明未来产量将经历更大的波动。这可能与气候变化有关，因为气候变化引起的极端气候事件，如干旱、洪涝等，会对农作物产量造成影响，从而增加产量的变异性。此外，本研究发现未来受 ENSO、IOD 和 NAO 显著影响的区域都将会增加（图 6-5c）。这可能意味着未来 ENSO、IOD 和 NAO 将会对粮食生产造成更明显的影响。然而，不同模型的结果表明，在大气环流指数对作物产量影响的分析中仍存在较大的不确定性。未来显著性影响的面积差异比较大，这表明需要更深入的研究来了解大气环流指数对作物生产的影响。

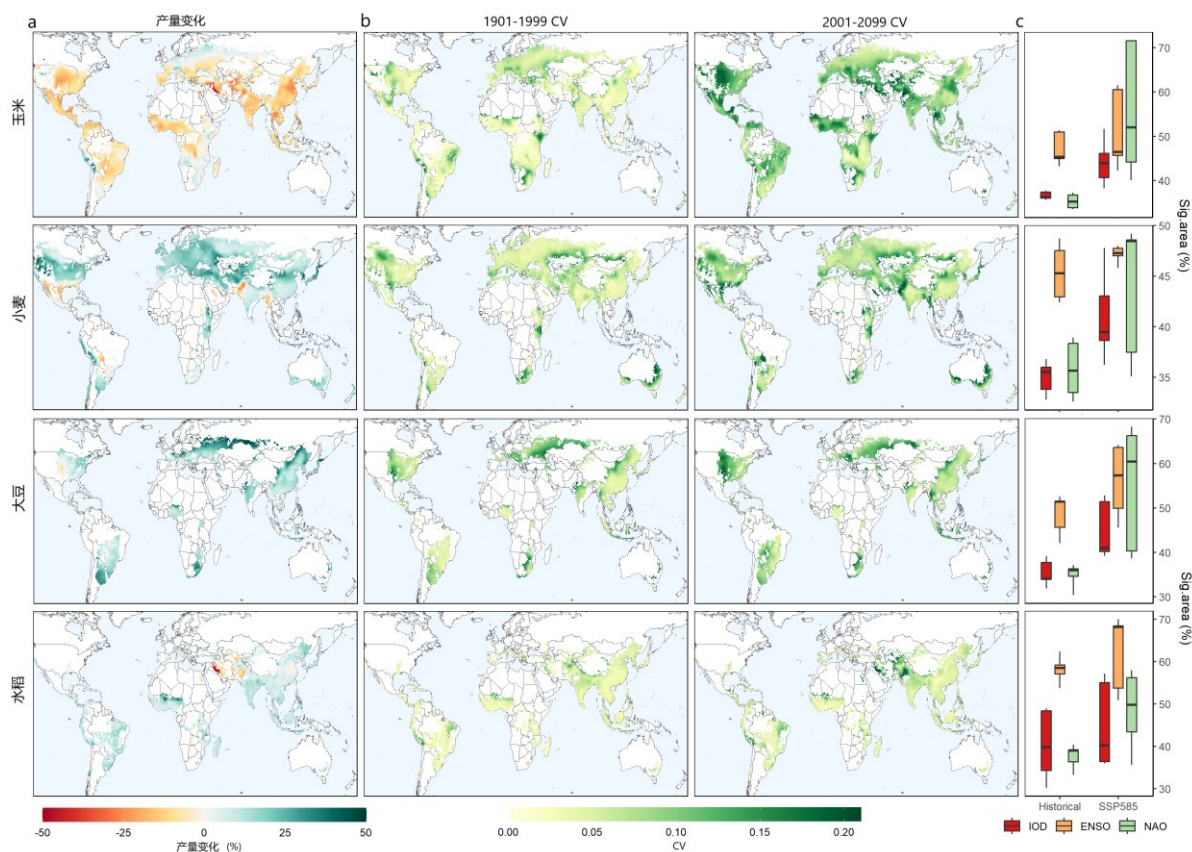


图 6-5 气候变化下作物产量变化。**a** 在 SSP585 下 1900-1999 年和 **b** 2000-2099 年作物产量变化和变异系数(CV)。**c** 受 ENSO、IOD 和 NAO 显著影响的种植区域

Fig. 6-5 Crop yield changes under climate change. **a** Crop yield changes and **b** coefficient of variation (CV) during 1900-1999 and 2000-2099 under SSP585. **c** Harvest areas that are significantly impacted by ENSO, IOD, and NAO.

### 6.2.3 主导因子的变化

通过机器学习得到了历史和未来主导因子的变化（图 6-6 和图 6-7），研究的结果显示了在基于 GFDL-ESM4 的 SSP126 和 SSP585 下的两个时期（历史和未来），每个网格和子区域（图 6-1）的主要气候驱动因素。图中（图 6-6，以 GFDL-ESM4 为例）可以看出对于小麦和玉米来说，在历史期间（1901-1999）ENSO 是产量变化主导因子，主要集中在澳洲、南美洲以及东南亚地区。在北美地区 IOD 占主导。然而在未来（2000-2099）期间，NAO 所占的比例不断增大，尤其是北半球的玉米。NAO 和 IOD 对欧洲和亚洲北部，以及北美洲地区小麦和玉米影响增大。南半球的玉米 ENSO 所影响的面积进一步增加，而小麦仍然由 IOD 和 ENSO 主导，且 IOD 有所增加。例如，在 1900-1999 年期间，小麦产量主要受 ENSO 在 OCE 中的影响，但预计 IOD 对 OCE 小麦产量影响的区域将在 2000-2099 年期间（图 6-6）有所增加。这一结果与 Feng 等（2022b）相似。本研究关注两个 99 年（1900-1999 年和 2000-2099 年），因为更长的时间可以减少内部变异的影响（Cai et al. 2022），并可以更好地揭示气候变化下气候振荡对作物产量的影响。主要气候驱动因素在这两个时期有很大的差异。

总的来说，在历史时期，ENSO 是作物产量的最重要的影响因素，占据了全球玉米、小麦、大豆和水稻种植面积的 26.7%、29.5%、32.7%和 43.7%。特别是在北美、亚洲和澳洲（见图 6-6 和图 6-7）。IOD 则是影响 CAN 和 USA 中玉米和大豆的主要因素。NAO 主要影响 CHN 和 EU 中的玉米、大豆和水稻产量。然而，预计在 2000-2099 年期间，NAO 影响区域将有所增加，尤其是在北半球（见图 6-6 和图 6-7）。此外，预计 BRA、OCE、SAF 的小麦受 IOD 影响区域也会增加。虽然在 SSP126 情景下 2000-2099 年各气候驱动因素主要影响的总体区域与历史相似，但在几个子区域仍有较大变化（见图 6-8）。例如，历史期间一些玉米和大豆一些地区不受大气环流显著影响，而未来则受 ENSO 显著影响（图 6-8）。

对于不同的 GCM，发现不同气候模式之间差异也比较大，例如，MRI-ESM2-0 和 MPI-ESM1-2-HR 无论历史还是未来，ENSO 仍然占主导因素且其影响的面积未来可能会变大。对于 UKESM1-0-L 和 IPSL-CM6A-LR 在 SSP126 情景下 ENSO 影响的区域最大。然而，在 SSP585 情景下，大部分地区的产量由 NAO 变为了主导，尤其是北半球的玉米和大豆。这可能是由于这两个气候模式是高平衡气候敏感性的模型，在未来 SSP585 下显示出更大的变暖，进而导致产量与其它模型差异较大。例如，UKESM1-0-LL 改进了中纬度混合相云的过程。这一过程导致平衡气候敏感性的增加，并显示出更大的未来变暖（Galik 2019）。

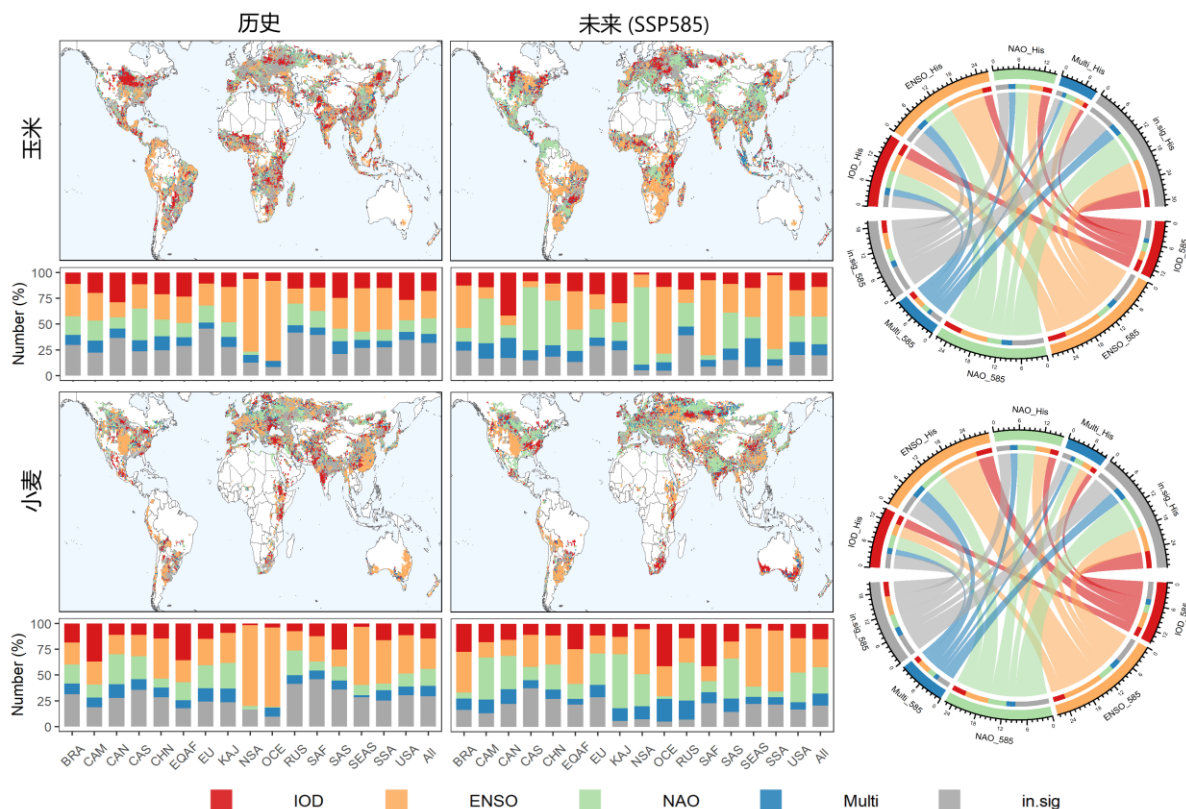


图 6-6 由 RF 模型确定的全球每个网格的玉米和小麦产量的主导气候振荡指数。a 和 c (顶部), SSP585 下历史 (1900-1999 年) 和未来 (2000-2099 年) 主导因素的空间分布。Multi, 网格受到不同指数的显著影响; in.sig, 无显著性指标。a 和 c (底部), 每个子区域中主要因素的面积比例。

和弦图显示了主导因素从历史 (1900-1999) 到未来 (2000-2099) 的变化区域

Fig. 6-6 Dominant climate oscillation indices of global maize and wheat yield at each grid as identified by the RF model based on GFDL-ESM4. **a** and **c** (top panel), the spatial distribution of dominant factors in historical (1900-1999) and future (2000-2099) under SSP585. Multi, the grid is significantly impacted by different indices; in.sig, no significant indices. **a** and **c** (bottom panel), the proportion of areas for the dominant factor in each sub-region. **b** and **d**, the chord diagram shows the shift areas in the dominant factor from historical (1900-1999) to future (2000-2099)

对于水稻和大豆来说也有类似的规律, 历史和未来相比, NAO 同样有着明显的提升 (图 6-7)。比如在历史期间 ENSO 为欧洲大豆产量变化的主导因子, 而在未来期间 (SSP585) 逐渐由 NAO 为主导。同样水稻也有类似的规律, 例如在东亚地区, 好多原来由 IOD 主导的区域逐渐被 NAO 替代。在南半球, ENSO 对水稻和大豆影响的面积有明显的增加。从和弦图中也可以看出 (图 6-6 和图 6-7), 全球大气环流指数对不同作物的影响格局发生了一个明显的转变。

本章以 GFDL-ESM4 为例介绍了未来主导因子的变化。其它气候模式的结果见附图 1 至附图 4。



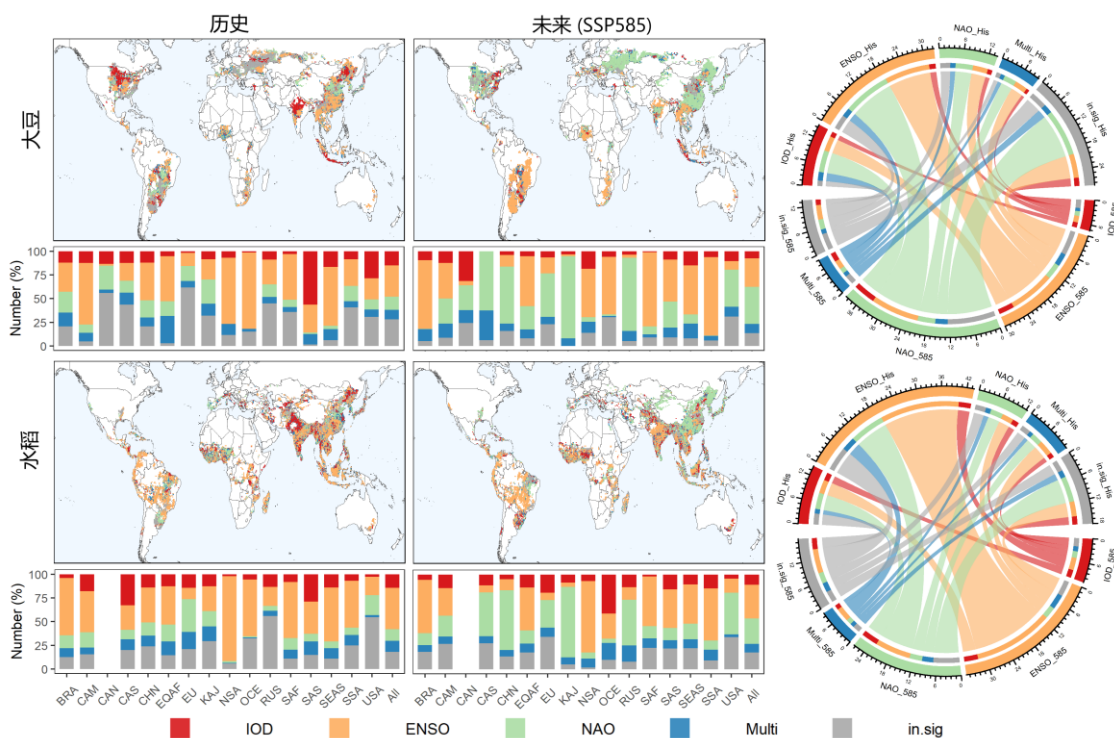


图 6-7 由 RF 模型确定的全球每个网格的大豆和水稻产量的主导气候振荡指数。a 和 c（顶部），SSP585 下历史（1900-1999 年）和未来（2000-2099 年）主导因素的空间分布。Multi，网格受到不同指数的显著影响；in.sig，无显著性指标。a 和 c（底部），每个子区域中主要因素的面积比例。

和弦图显示了主导因素从历史（1900-1999）到未来（2000-2099）的变化区域

Fig. 6-7 Dominant climate oscillation indices of global soybean and rice yield at each grid as identified by the RF model based on GFDL-ESM4. **a** and **c** (top panel), the spatial distribution of dominant factors in historical (1900-1999) and future (2000-2099) under SSP585. Multi, the grid is significantly impacted by different indices; in.sig, no significant indices. **a** and **c** (bottom panel), the proportion of areas for the dominant factor in each sub-region. **b** and **d**, the chord diagram shows the shift areas in the dominant factor from historical (1900-1999) to future (2000-2099)

尽管在 SSP585 排放情境下 NAO 成为主导因子，但在 SSP126 中，NAO 的影响并不如此明显（见图 6-8）。总体来看，不显著的区域正在减少，这意味着更多的地区的作物产量可能会受到大气环流的影响。本研究进行了对比，分析了产量去趋势后的结果，以确定这种变化是否是由于升温或二氧化碳浓度上升所导致的。结果表明，在去趋势的情况下，NAO 的影响区域并没有明显增加，而 ENSO 仍然是大多数区域作物产量的主要影响因素（见图 6-9）。值得注意的是，对于四种作物，ENSO 对产量的影响区域仍在增加（无论是在 SSP126 还是 SSP585 情境下），这表明即使不考虑全球变暖等因素，ENSO 的影响仍在加强。然而，需要注意的是，本研究没有着重分析对产量去趋势的结果，因为作物模型的模拟结果是不包含管理措施和品种改良的。因此，图 6-9 仅用于对比分析，以反映仅考虑气候因素的情况。

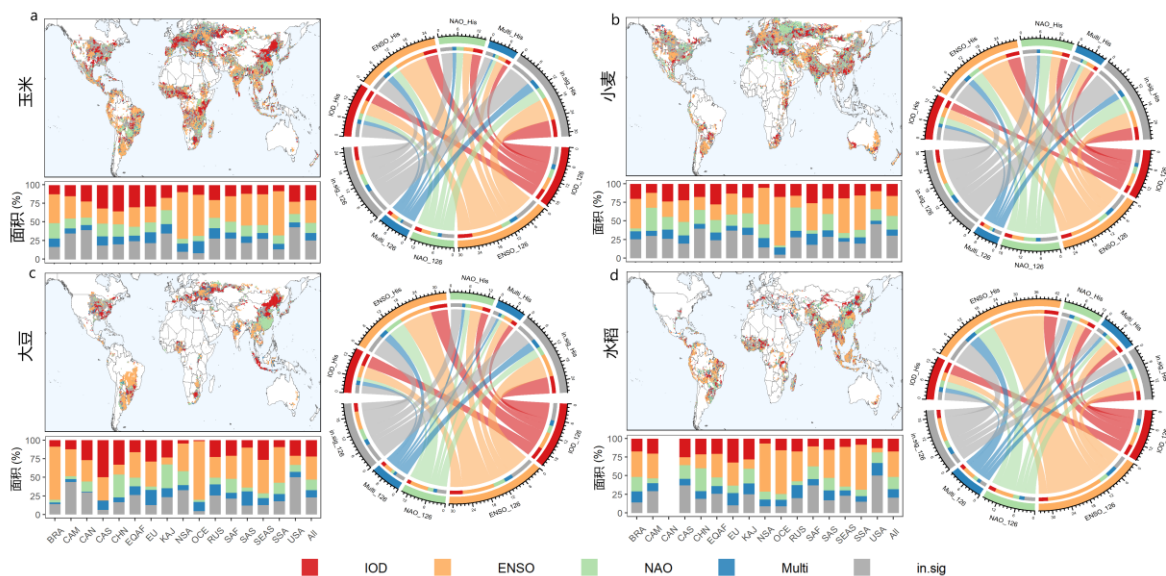


图 6-8 基于 MRI-ESM2-0 气候模式，由 RF 模型确定的每个网格的玉米、小麦、大豆和水稻在 SSP126 情景下产量的主导气候振荡指数。

Fig. 6-8 Dominant climate oscillation indices of global maize, wheat, soybean and rice yield at each grid as identified by the RF model based on GFDL-ESM4 under SSP126.

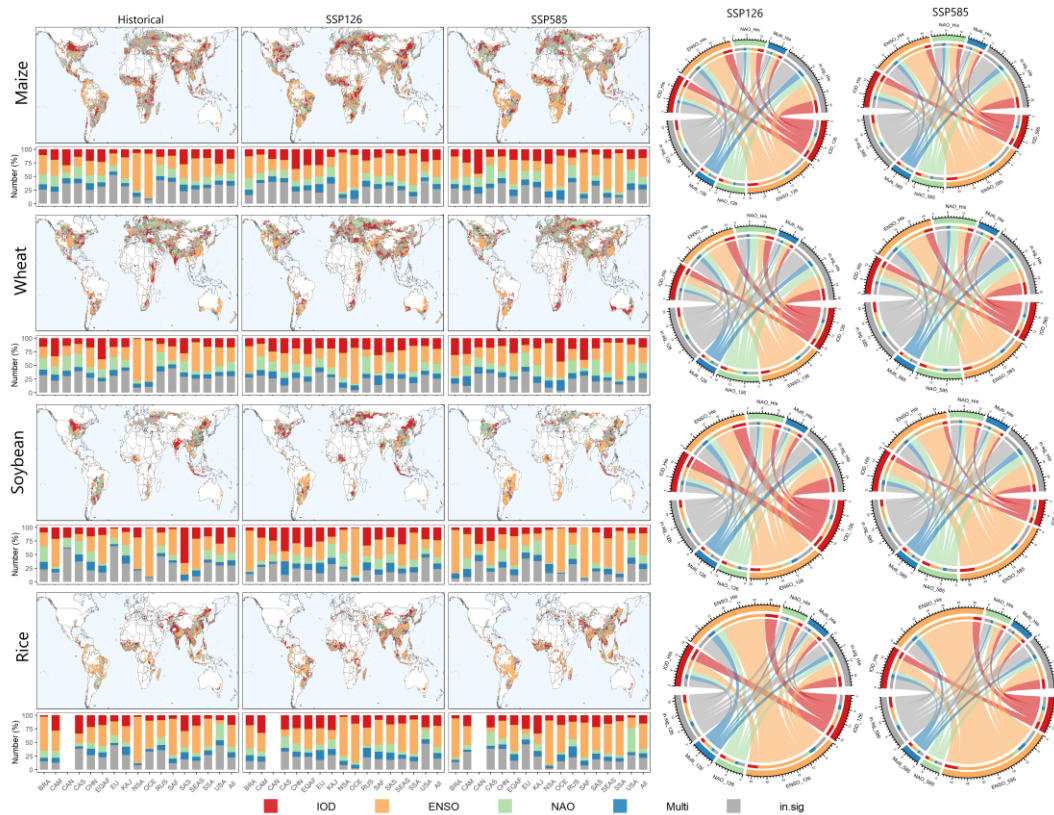


图 6-9 基于 GFDL-ESM4 气候模式，由 RF 模型确定的每个网格的玉米、小麦、大豆和水稻产量（去趋势）的主导大气环流指数

Fig. 6-9 Dominant climate oscillation indices of global maize, wheat, soybean and rice yield at each grid as identified by the RF model based on GFDL-ESM4

### 6.2.4 强烈震荡阶段产量的变化

本研究通过随机森林的局部依赖函数（PDP），量化了在大气环流指数强烈震荡阶段对产量的影响。通过找出该时段因子 75 分位数（25 分位数）作为强烈震荡，然后通过局部依赖关系曲线来量化它们的产量。通过将强烈震荡期间的产量减去正常（指数在 50 分位数）的产量，来计算产量变化。图 6-10 展示了以一个网格为例，大气环流因子与产量的 PDP 关系曲线。红色和蓝色的线分别表示 25 分位数和 75 分位数的值。本研究用同样的方法对每个网格进行识别，并量化它们的产量。

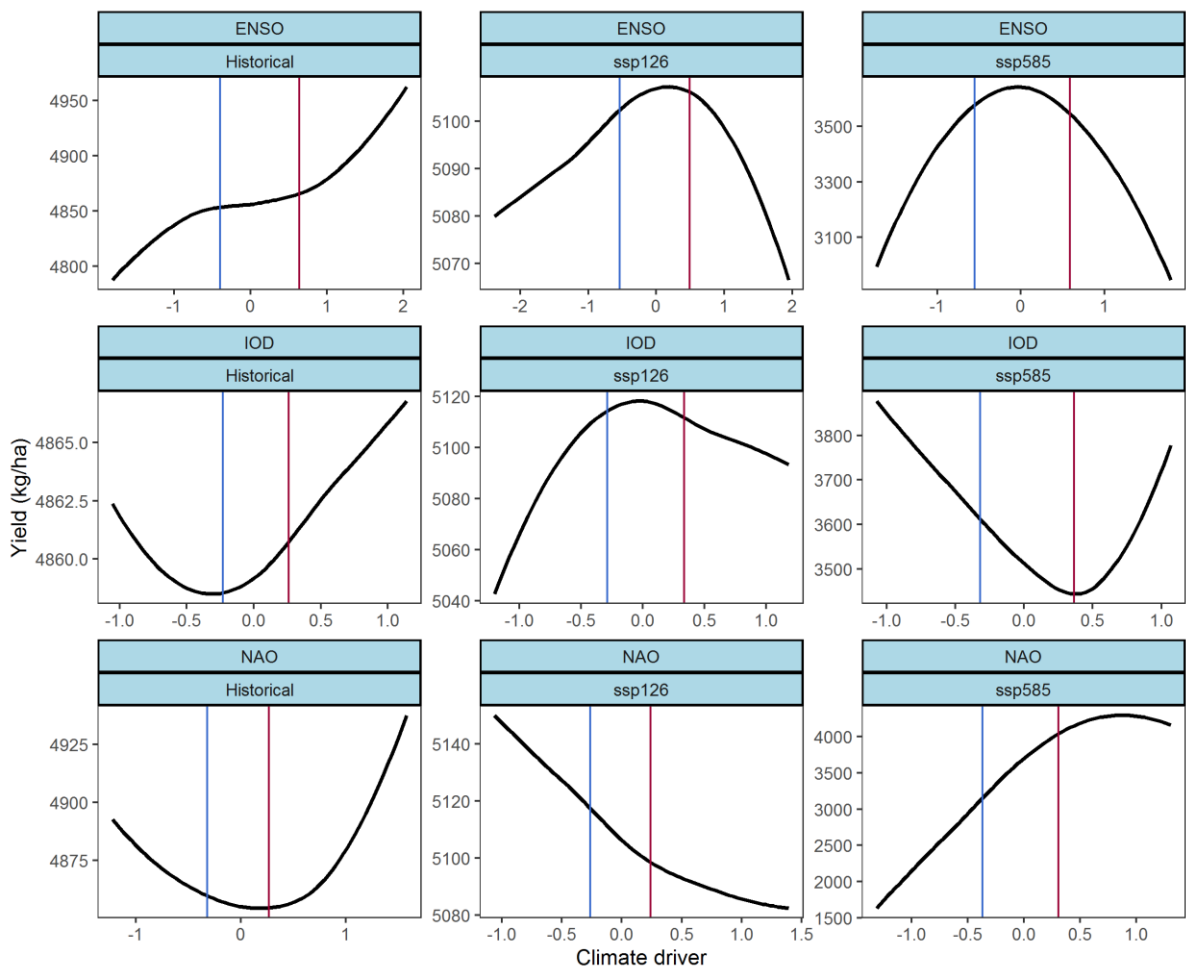


图 6-10 基于 GFDL-ESM4 气候模式，由局部依赖关系确定的大气环流强烈震荡对每个网格的小麦产量的影响。箱型图为不同区域产量变化

Fig. 6-10 Wheat yield change during strong oscillation phases quantified by partial dependence plot based on GFDL-ESM4 under SSP585 (top panel). The box plot shows different sub-regions yield change under strong climate oscillation phases



图 6-11 和图 6-12 展示了在强震荡阶段作物产量的变化情况。在 SSP585 情景下，ENSO 和 NAO 的负相位异常导致了作物产量的剧烈波动，尤其是在北半球，减产现象更为显著。另一方面，在历史时期，南半球的某些地区在 ENSO 负相位时可能会出现一定程度的增产，但在未来这一增产效果将会显著减弱。对于 ENSO 正相位异常（厄尔尼诺现象）而言，未来其对北半球的负面影响可能会减小，然而对于南半球（如澳大利亚）来说，可能会引发更严重的减产现象。IOD 异常对北美、欧洲和澳大利亚等地区产生较大影响。在负相位异常下，澳大利亚和亚洲南部等地区的产量可能会增加，而正相位异常可能导致北欧和北美地区的产量减少。总体来看，在空间分布上，IOD 异常在历史时期和未来表现出相似的规律，但在未来的 SSP585 情景下，变化幅度可能会进一步加大。

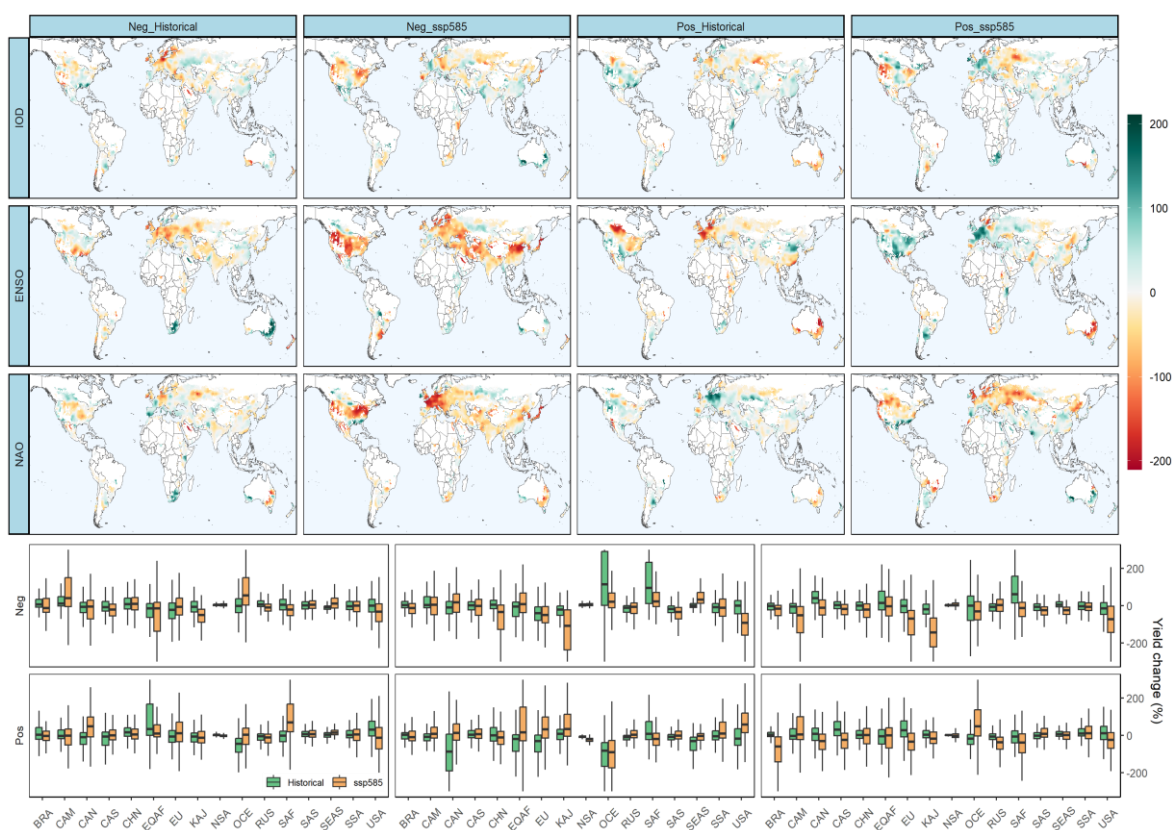


图 6-11 基于 GFDL-ESM4 气候模式 SSP585 下历史和未来，由局部依赖关系确定的大气环流强烈震荡对每个网格的小麦产量的影响。箱型图为不同区域产量变化

Fig. 6-11 Wheat yield change during strong oscillation phases quantified by partial dependence plot based on GFDL-ESM4 under SSP585 (including historical and future). The box plot shows different sub-regions yield change under strong climate oscillation phases

在 SSP126 情景下（图 6-12），作物产量的变化幅度没有 SSP585 情景下的显著，但与历史时期相比，整体影响存在较大差异。例如，在历史时期，IOD 异常可能导致

欧洲地区减产，然而在未来 SSP126 情景下，IOD 负相位异常可能会促使北欧地区的产量略有增加。此外，ENSO 负相位（拉尼娜现象）对作物产量的负面影响可能会减轻。而 NAO 异常对作物产量影响的分布在未来与历史时期的总体趋势相似。

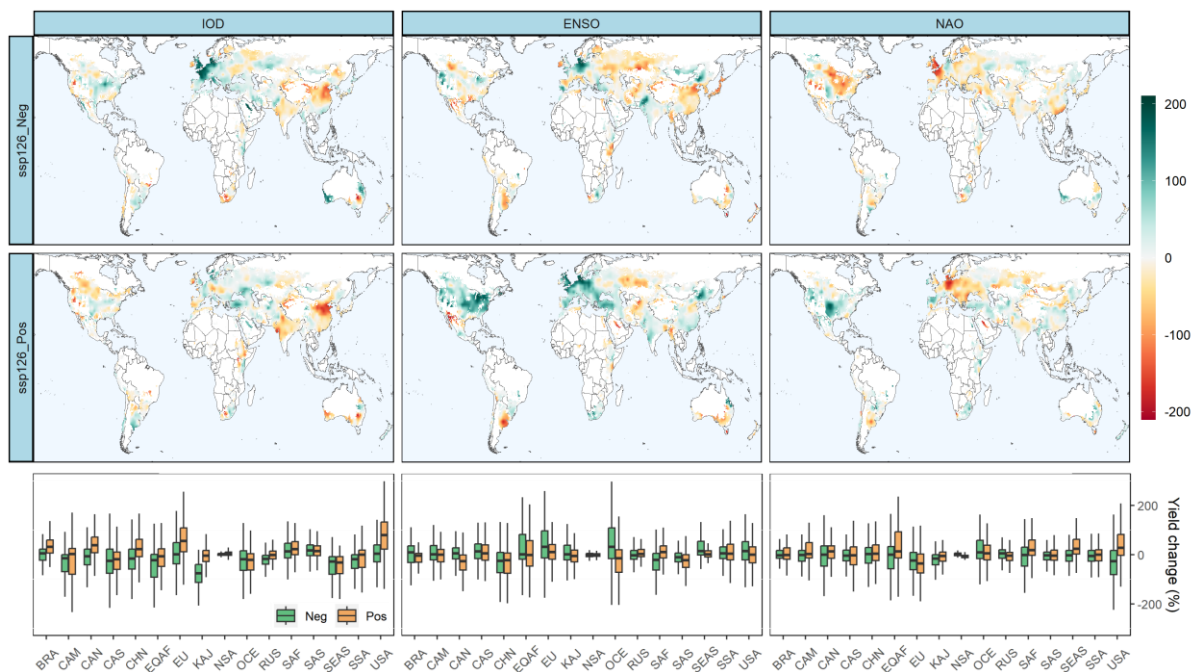


图 6-12 基于 GFDL-ESM4 气候模式 SSP126 情景，由局部依赖关系确定的大气环流强烈震荡对每个网格的小麦产量的影响。箱型图为不同区域产量变化

Fig. 6-12 Wheat yield change during strong oscillation phases quantified by partial dependence plot based on GFDL-ESM4 under SSP126 (future: 2000-2099). The box plot shows different sub-regions yield change under strong climate oscillation phases

上述介绍了小麦产量变化。在模拟的不同气候模式下，对于其它不同作物，产量变化会受到不同程度的影响（见附图 5-9）。然而，我们能否从这些不确定性的结果中梳理出一些共性呢？为了识别哪些指数异常会对全球粮食安全造成威胁，本研究统计了在强烈震荡阶段产量显著变化的面积（见图 6-13）在 SSP585 情景下，NAO 负向异常会导致玉米、小麦、水稻大面积减产，而 NAO 正向异常将导致玉米和大豆增产。未来 ENSO 负向异常（拉尼娜）将促进玉米和大豆大面积产量增加，而正向异常（厄尔尼诺）将导致玉米、大豆、水稻大面积减产。关注这些指数的异常有利于更早地感知全球粮食安全风险。附图 10-13 展示了不同子区域在强震荡阶段的产量变化。例如，在 SSP585 情景下，IOD 负向的异常会导致 USA、EU 和 CAN 等地区减产，但是会导致 SEAS、BRA 和 CAM 等地区增产。

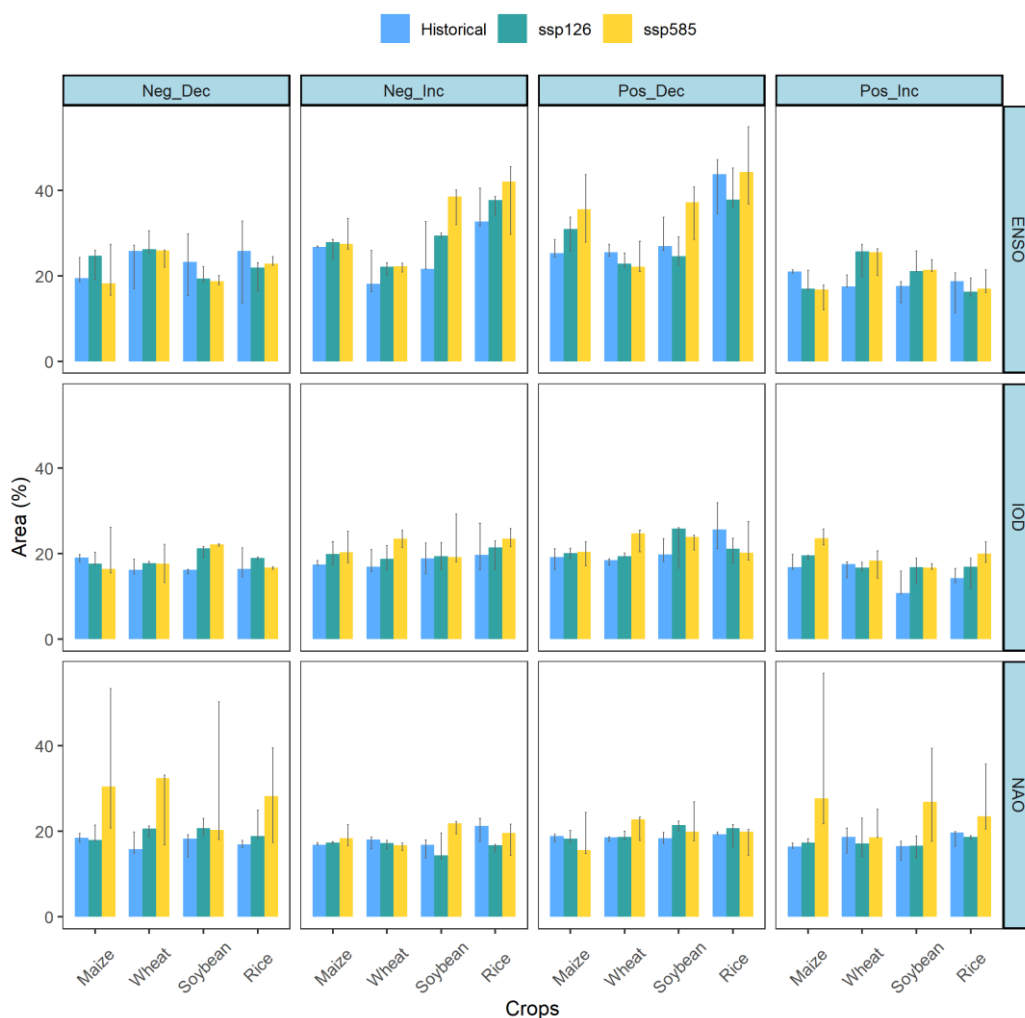


图 6-13 基于历史和未来 5 个 GCMs(SSP126 和 SSP585),农作物产量在强振荡阶段的显著变化区域。误差条表示 5 个 GCM 的标准误差

Fig. 6-13 Crop yield significant change areas during strong oscillation phases based on 5 GCMs in the historical and future (SSP126 and SSP585). The error bars represent the standard errors for 5 GCMs

### 6.2.5 产量对大气环流指数的敏感性

使用动态线性模型 (DLM) 分析了作物产量对 IOD、ENSO 和 NAO 的敏感性变化。对于玉米来说,与历史相比,SSP585 情景下未来产量对 ENSO、IOD 和 NAO 的敏感性都有所增加。历史期间,相对于 ENSO 和 NAO,IOD 与玉米产量更为敏感(图 6-14)。在未来,NAO 与北半球的玉米产量有强正相关关系。对于小麦来说,与历史相比,IOD 的敏感性有所增加(图 6-14),但整体的关系没有发生变化。在北半球,ENSO 与小麦产量呈正相关关系,在南半球(尤其是澳大利亚小麦带)则呈负相关。NAO 的历史和未来趋势相似,与美国东部、欧洲西南部、亚洲和澳洲的小麦产量呈正相关关系,而与其它地区呈负相关关系。在未来,NAO 的敏感性也有所增加。

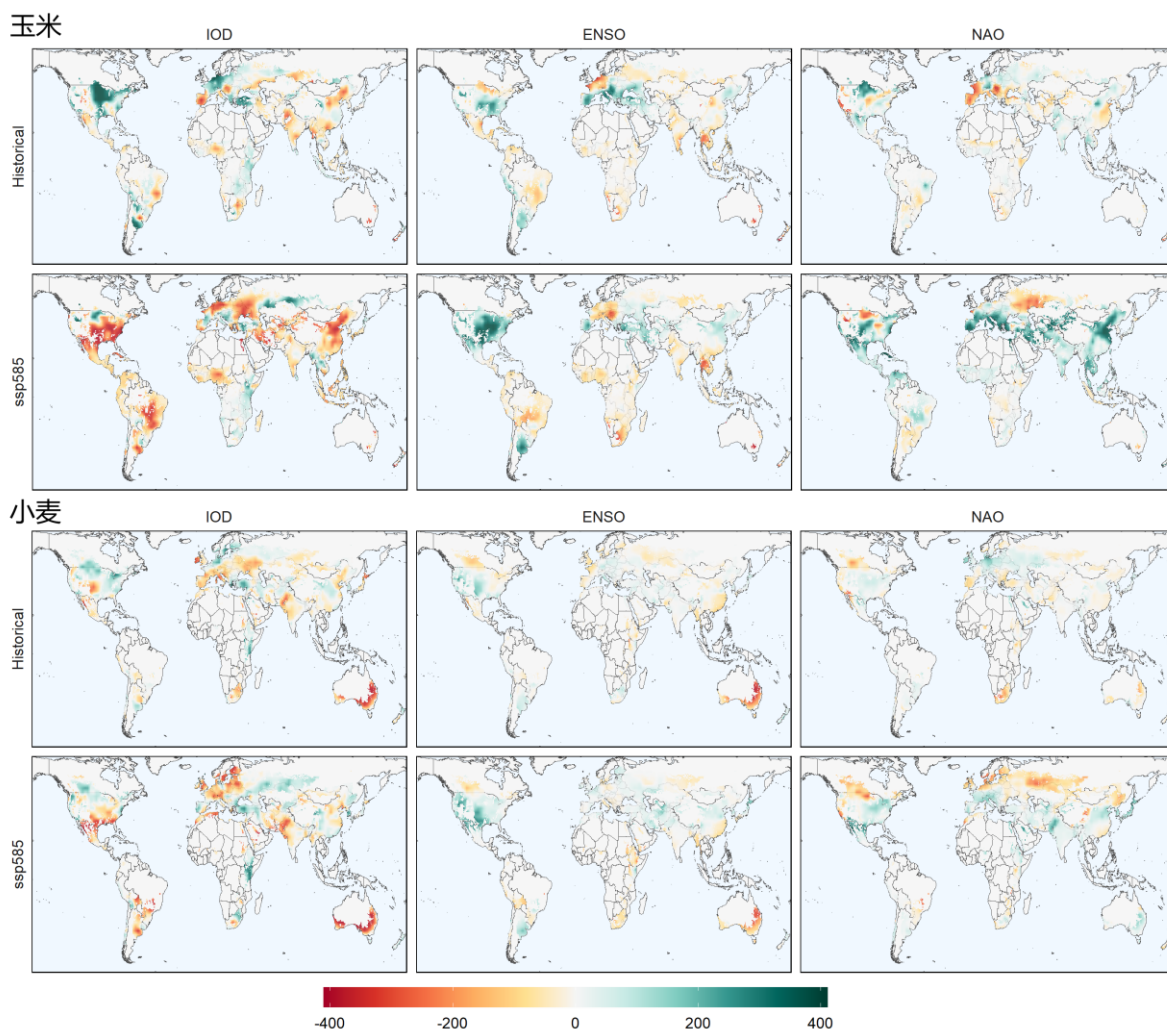


图 6-14 历史(1900-1999)和未来(SSP585, 2000-2099)玉米和小麦产量对 IOD、ENSO 和 NAO 的敏感性。敏感性值由基于 GFDL-ESM4 的 DLM 计算。

Fig. 6-14 Maize and wheat yield sensitivity to IOD, ENSO, and NAO during historical (1900-1999) and future (SSP585, 2000-2099). The sensitivity values are calculated by DLM based on GFDL-ESM4.

对于大豆来说, IOD 在北美地区由正相关变成了负相关, NAO 在北半球有正相关, 未来变为了明显的负相关。而对于水稻来说, IOD 历史和未来变化不大, ENSO 空间分布比较接近, 未来敏感性有略微提高。与大豆一样, NAO 在北半球对水稻产量有很强的负相关, 尤其是在未来期间(图 6-15)。

总体来看, 未来产量对气候指数的敏感性有所提高, 尤其是对于大豆和水稻来说, NAO 对北半球产量有很强的负相关。



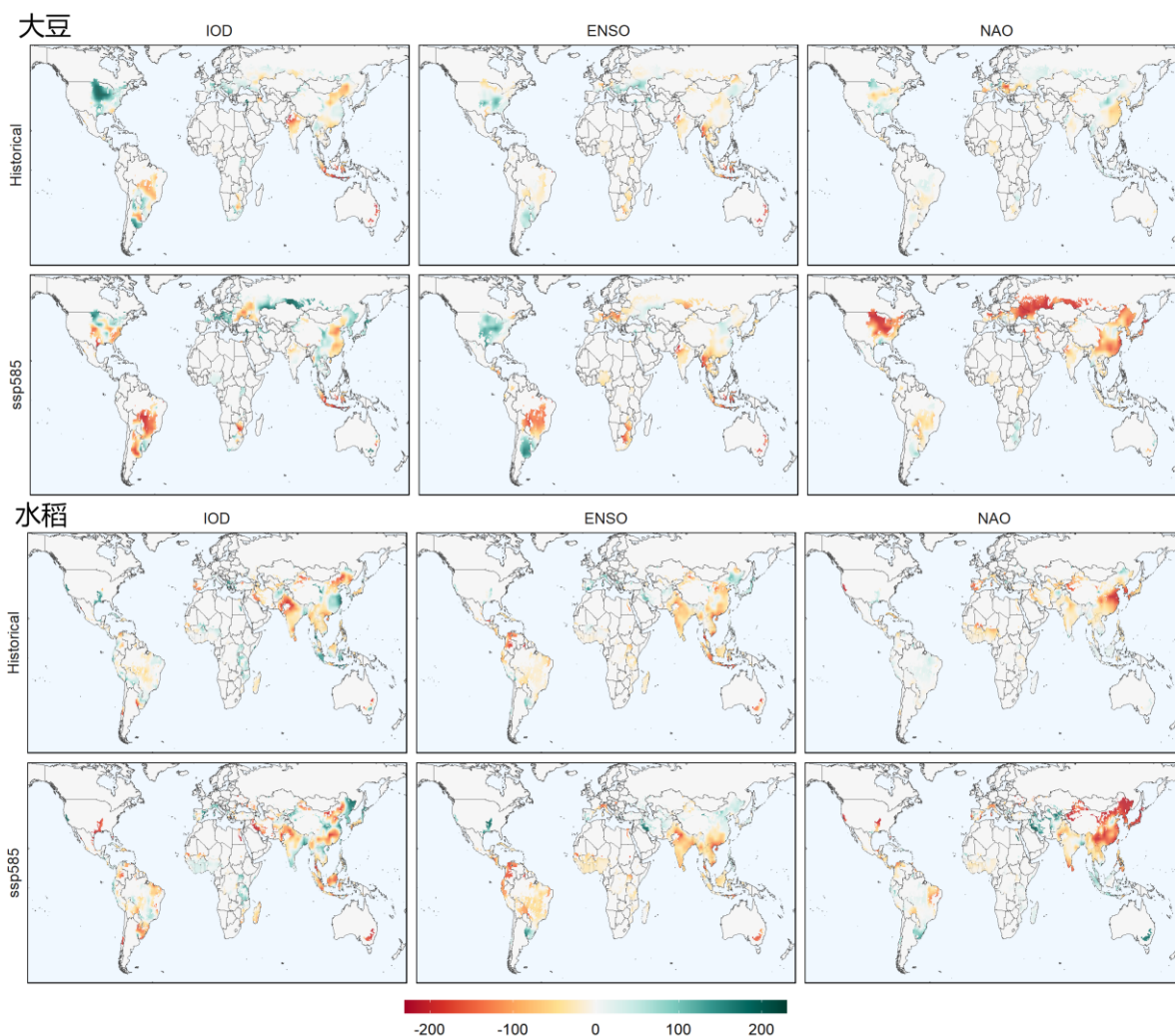


图 6-15 历史(1900-1999)和未来(SSP585, 2000-2099)大豆和水稻产量对 IOD、ENSO 和 NAO 的敏感性。敏感性值由基于 GFDL-ESM4 的 DLM 计算。

Fig. 6-15 Soybean and rice yield sensitivity to IOD, ENSO, and NAO during historical (1900-1999) and future (SSP585, 2000-2099). The sensitivity values are calculated by DLM based on GFDL-ESM4.

对于 SSP126 来说, 整体上与历史期间一致, 但是玉米对 IOD、ENSO 和 NAO 的敏感性比其它作物要高。小麦产量对 IOD 有较高的敏感性(图 6-16)。水稻和大豆对 IOD 比较敏感, 且空间分布比较相似。不同作物之间的差异可能与管理措施, 生育期, 作物品种等因素有关系。

这部分只展示了一个 GCM 的结果, 对于不同 GCM 来说它们之间存在较大差别。例如在 SSP585 情境下, GFDL-ESM4 预测的 ENSO 未来与北美地区的玉米为负相关, 然而在 UKESM1-0-1-LL 预测中, ENSO 与该地区玉米产量为正相关。说明不同 GCM 之间可能存在很大的不确定性(附图 13 和附图 14)。

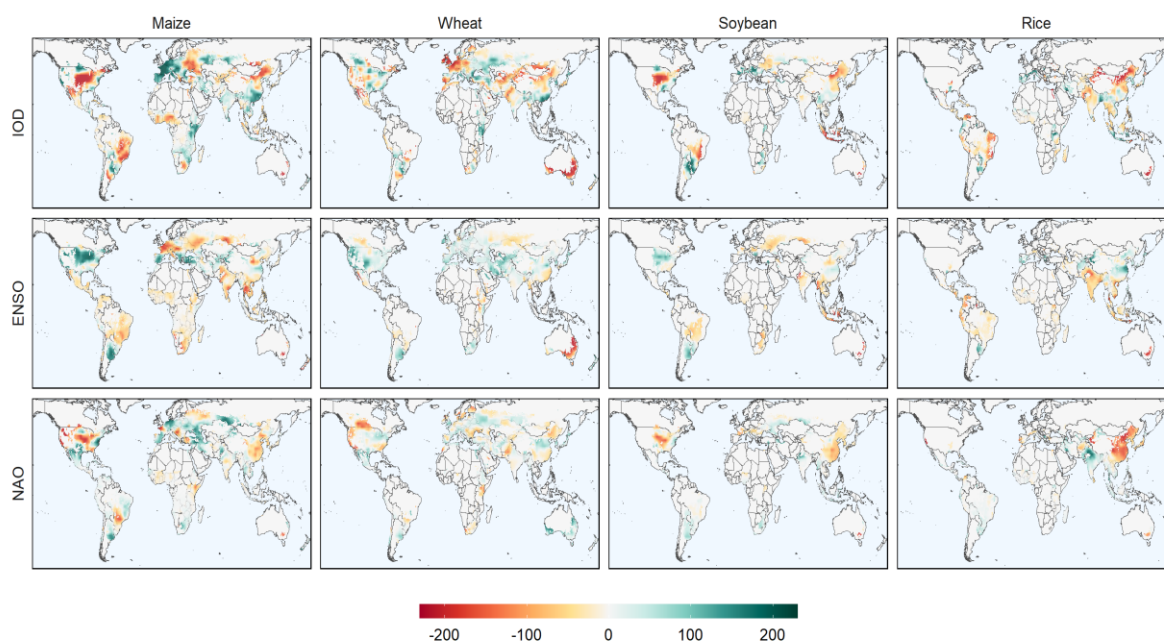


图 6-16 历史(1900-1999)和未来(SSP126, 2000-2099)玉米、小麦、大豆和水稻产量对 IOD、ENSO 和 NAO 的敏感性。敏感性值由基于 GFDL-ESM4 的 DLM 计算。

Fig. 6-16 Maize, Wheat, Soybean and rice yield sensitivity to IOD, ENSO, and NAO during historical (1900-1999) and future (SSP126, 2000-2099). The sensitivity values are calculated by DLM based on GFDL-ESM4

## 6.2.6 研究的不确定性分析

这个研究通过敏感性分析发现不同区域的不确定性来源因地而异。在大多数作物和指标中，GCM 和 SSP 以及它们之间的交互往往是不确定性的主要来源，贡献超过 60%。总体而言，GCM 的不确定性贡献比 SSP 更大，而 SSP 的不确定性贡献则大于 GGCM（图 6-17 和图 6-18）。在水稻和玉米产量对 NAO 的敏感性分析中，SSP 的不确定性贡献高于 GCM，这表明在分析大气环流场和产量之间的关系时，GCM 是不确定性的主要来源。这可能是因为在分析气候驱动因子与未来作物产量的关系时，气候模型的不确定性会变得更加重要。因为在考虑未来环流因子的相位的时候，气候模式所造成的不确定性会被放大。

另一方面，在预测未来作物产量时，作物模型往往是主要的不确定性来源。这是因为作物模型的预测受到多种因素的影响，包括管理措施、生育期、作物品种等等，而这些因素在未来很难准确预测。然而，当再分析大气环流因子（如 ENSO、IOD 和 NAO）与未来产量的关系时，气候模型的不确定性又变得更加重要。这是因为本研究预测的大气环流因子是基于气候模型数据计算的，因此在未来情境下，气候模型的不确定性将进一步影响大气环流与产量关系的预测结果，进而成为不确定性的主要来源。

鉴于作物模型在本研究中所占的不确定性很小，因此，在分析本章时，把作物模



型的结果取平均，然后单独考虑不同的气候模式。

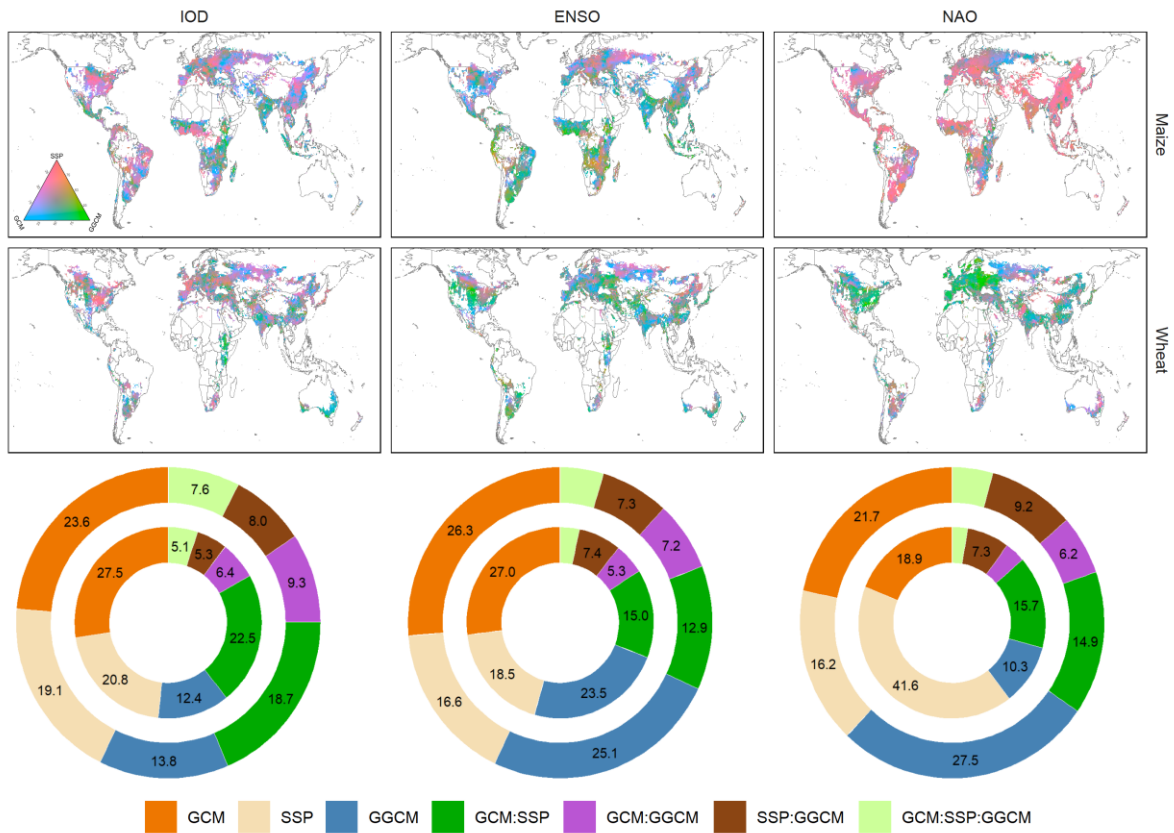


图 6-17 玉米和小麦的 GCM、GGCM 和 SSP 敏感值不确定度的地理分布(上面板)。SSP 相关不确定性包括两个时期(1900-1999 年和 2000-2099 年)。饼图显示了敏感性的值的不确定度占比。内圈代表玉米的不确定性份额，外圈代表小麦的不确定性份额。

Fig. 6-17 Geographic distribution of the uncertainty of sensitivity value sourced from GCM, GGCM, and SSP of maize and wheat (top panel). The SSP associates uncertainty include two periods (1900-1999 and 2000-2099). The pie chart shows the proportion of uncertainty of sensitive value. The inner ring represents the uncertainty share of maize, and the outer circle represents the uncertainty share of wheat

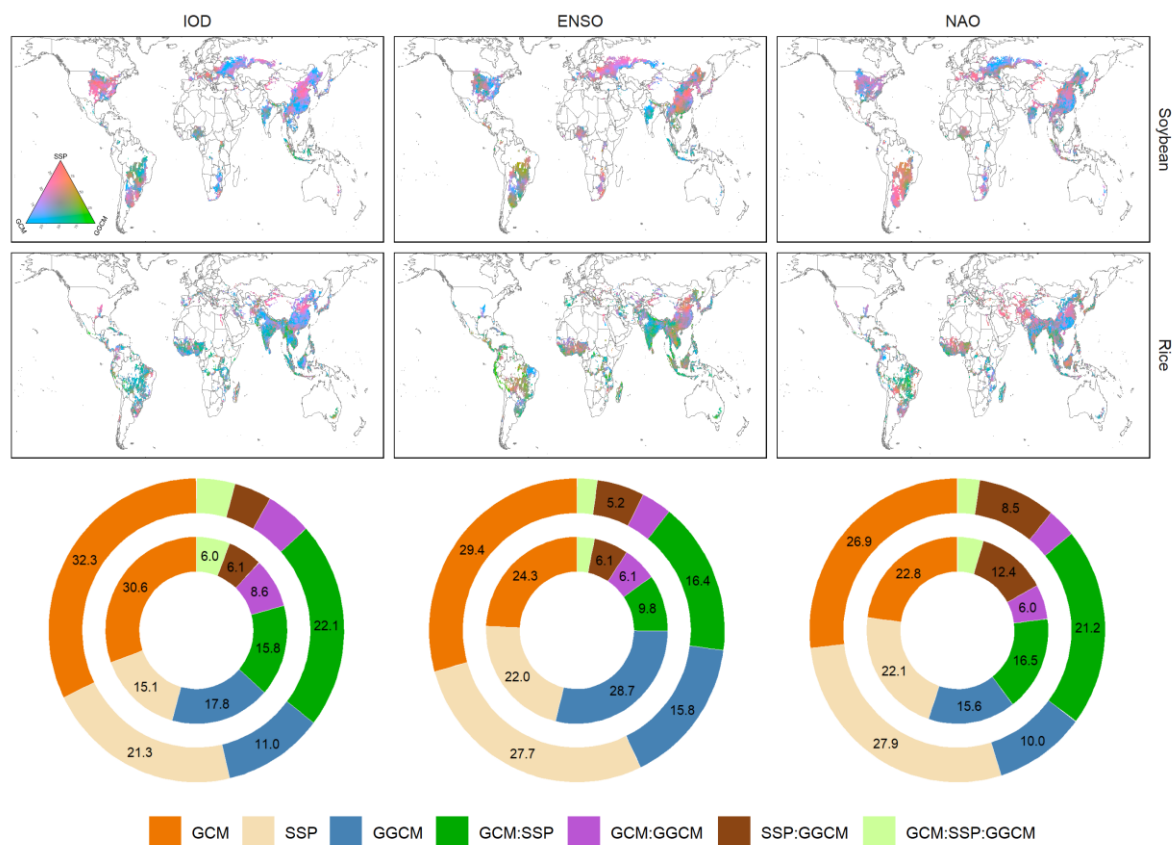


图 6-18 大豆和水稻的 GCM、GGCM 和 SSP 敏感值不确定度的地理分布(上面板)。SSP 相关不确定性包括两个时期(1900-1999 年和 2000-2099 年)。饼图中显示了敏感性的值的不确定度占比。内圈代表大豆的不确定性份额，外圈代表玉米的不确定性份额。

Fig. 6-18 Geographic distribution of the uncertainty of sensitivity value sourced from GCM, GGCM, and SSP of soybean and rice (top panel). The SSP associates uncertainty include two periods (1900-1999 and 2000-2099). The pie chart shows the proportion of uncertainty of sensitive value. The inner ring represents the uncertainty share of soybean, and the outer circle represents the uncertainty share of rice

## 6.3 讨论

### 6.3.1 气候变化导致对产量影响的主导因子发生了转变

本研究结果显示，对于玉米（SSP585）而言，欧洲地区、中亚地区和美国西部地区是未来变化最为明显的区域，而在历史时期中美国西部地区并未受到显著影响因素的影响，但未来变北大西洋涛动（NAO）将成为该地区的主导影响因子（如图 6-6 所示）。同时，欧洲和中亚地区的主导因子也将从历史时期的 ENSO 转变为 NAO。值得注意的是，这种变化在不同气候模型的预测中得到了一致的验证。NAO 是一个海平面压力偶极子，它控制着北美和欧亚大陆的急流位置和冬季环流模式，与北半球极端气候的形成密切相关（Nilsson and Lejenäs 2011）。另外，由于全球变暖可能导致北半

球经向温度梯度降低，中纬度夏季环流减弱，这有助于持续的热浪和大气阻塞，并进一步影响作物产量 (Sippel et al. 2018)。在不同 GCM 的预测中，气候变化将导致 NAO 对欧洲等地区的降水起着更为重要的作用 (McKenna and Maycock 2022)。在气候模式模拟的高排放情景中，ENSO 和 NAO 的遥相关增强，进而增强了热带降雨与 NAO 的关系 (Fereday et al. 2020)。这将导致一些中纬度地区产量与 NAO 之间的关系增强。在 SSP585 情景下，IOD 对一些小麦种植区域的影响也将增强，例如北美地区和澳大利亚地区将从 ENSO 主导转变为 IOD 主导。这可能是由于气候变化增强了极端 IOD 事件的发生，而在 IOD 正相位（即“pIOD”），苏门答腊海面温度（SST）异常偏冷，而在北半球夏季和秋季，西印度洋上空的海温异常偏暖，从而导致严重的降水（洪水）和东非、印度尼西亚的干旱 (Wang et al. 2021b)，从而进一步导致产量减少。对于大豆和水稻，可以看到它们也有类似的规律，即 NAO 所主导的区域有明显的增加（图 6-7）。

在 SSP126 排放情景下，整体变化没有 SSP585 的明显，但 NAO 对玉米、大豆和水稻的影响都是呈现增加的趋势。玉米整体与历史比较相似，但是在很多地区仍然出现一部分区域由无显著的影响因子变为由 IOD 和 ENSO 主导。IOD 对小麦、大豆和水稻的影响增加都由所增加。ENSO 对玉米的影响增加，但是对小麦水稻和大豆有所下降。对比 SSP585 与 SSP126 可以看出升温可能导致大气环流因子对产量的影响增加。

然而，可否把这个变化归因于 CO<sub>2</sub> 与温度升高呢？通过对产量进行去趋势处理来抵消这些影响，只关注大气环流因子不同阶段对产量的影响。本研究发现，未来仍然是 ENSO 主导作物产量变化，这表明二氧化碳浓度升高后，作物对 NAO 的响应更加敏感。除此之外，其它的响应也有所变化。例如，当不考虑气候变化中二氧化碳浓度以及升温等对产量的影响时，玉米和大豆北半球主导因子是 ENSO 和 IOD，而不是 NAO（见图 6-6 和图 6-7）。这说明，气候变化可能会导致 NAO 对北半球作物产量的影响增强，从而对粮食安全构成威胁。当然也有一些区域上差异不大，例如澳洲地区小麦在历史期间受 ENSO 主导，未来主要受 IOD 主导 (SSP585)。

对于不同气候模式，在历史期间和 SSP126 情景下，不同 GCM 表现的基本一致（附图 1-4）。然而在 SSP585 情景下，不同 GCM 之间差异较大。高气候平衡敏感性模型，如 UKESM1-0-L 和 IPSL-CM6A-L，对温度和二氧化碳浓度的响应更加敏感，因此模型中显示产量对 NAO 更加敏感。相比之下，中气候平衡敏感性的作物模型则假设作物对气候变化的响应相对较为缓和。因此，MPI-ESM1-2-HR 和 MRI-ESM2-0 的结果显示，ENSO 和 IOD 主导的区域更多，即使在 SSP585 情景下也是如此。这与之前做的去趋势处理的结果对比是一致的（图 6-9）。总体而言，升温可能导致一些指标（如 ENSO）的变异性增加 (Cai et al. 2022)。但是本研究说明，升温对大气环流

因子对产量影响分析仍然有较大的不确定性。

本研究识别出产量变化的主导因子，这对于农业生产有很大意义。首先，这些估算结果可以为农业生产提供预测信息，帮助农民和农业生产者更好地应对气候变化和环境波动的风险，调整农业生产计划和种植结构，以提高作物产量和质量，保障粮食安全和农民收入。例如，未来 ENSO 与作物产量的关系可能发生变化，一些地区可能会受到更大的影响。这就要求农业生产者要加强对气象变化的观测和预测，及时调整种植结构和农业生产计划，提高农业适应性和应对风险的能力。此外，确定主要气候驱动因素并了解其对作物产量和粮食生产的影响有助于促进粮食贸易。通过了解这些气候驱动因素如何影响不同地区的粮食生产，政策制定者和农民可以就粮食贸易做出明智的决定，例如从条件有利的地区增加粮食出口，或从其他地区进口粮食，以确保粮食短缺时期的粮食安全。了解气候驱动因素对粮食贸易的影响有助于降低风险，促进全球粮食市场的稳定。这在严重依赖少数主要作物的地区尤其重要，因为气候模式的变化可能对粮食生产和价格产生严重影响。通过为这些影响做好准备并就粮食贸易做出合适的决定，政策制定者可以帮助确保本国和世界各地的粮食安全和稳定。

### 6.3.2 大气环流强震荡对产量的影响

基于 PDP 图，可以量化出强震荡阶段它们对产量的影响。在历史期间不同方向的震荡往往导致产量也朝着相反的方向变化，这与过去的研究一致 (Heino et al. 2018)。然而在气候变化的影响下，相反的震荡可能都会导致作物减产或增产。例如 IOD 极端正和极端负向异常可能都会导致北美西部地区小麦产量下降 (图 6-11)。NAO 极端正向和极端负向异常都会导致欧洲和亚洲北部小麦产量下降 (SSP585)。在 SSP126 情景下，NAO 在负相位的时候又会导致欧洲和北美地区减产 (图 6-12)。与历史相比，未来气候变化可能会导致 ENSO、IOD 和 NAO 异常对产量影响增加，尤其是北半球的北美，欧洲以及东亚地区。不同气候模式下的预测的气候异常对产量影响整体比较一致，但是在一些地区也有区别，尤其是 ENSO 对小麦和大豆影响方面。例如在 IPSL-CM6A-LR 的预测中，SSP585 情景下，ENSO 正向异常导致北美地区增产，而在 MRI-ESM2-0 则是减产。说明在预测中还是存在一些不确定性。

那么，在这样的情况下，能否为全球粮食安全提供一些可靠的信息呢？通过统计 ENSO、IOD 和 NAO 异常对产量影响的面积，可以大致梳理出哪些指标异的常将导致全球多个粮食产区同时受到冲击。例如，研究发现 ENSO 在正相位异常时，玉米、大豆和水稻产量会有大面积减产。因此，在 ENSO 出现异常时应加强管理，以及为确保全球粮食安全做准备。此外，一些 ENSO-IOD 或 ENSO-NAO 等复合影响也需要注意。通过这样的分析可以为全球粮食安全提供有用的信息。例如，通过多样化种植，缓解厄尔尼诺为某一种作物带来的风险。假设在厄尔尼诺情况下，某地区小麦的产量

可能会下降,但玉米和大豆的产量可能会增加,因为 ENSO 对不同作物的影响方式是不同的。此外,可以加强农业气象服务系统,为农民提供准确的气象信息和农业技术指导,帮助农民更好地应对气候变化对农业生产的影响。

### 6.3.3 本研究的不确定性来源

通过对敏感性值的不确定性分析,发现本研究的主要不确定性来源是 GCM 以及 SSP。不同的温室气体排放情景对大气环流影响产量的方向有很大影响,这在对比两个排放情景中十分明显。尽管本研究中使用了 12 个作物模型和 5 个 GCM,但是 GCM 所占的不确定性仍然是主导因素。这似乎与其他研究预测产量不确定性的来源有所不同 (Müller et al. 2021; Xiong et al. 2019)。GCM 所造成的不确定性可能会被进一步放大,主要原因是在估产时要基于 GCM 来驱动物物模型,在分析大气环流因子对产量影响时,也要基于 GCM 来估算,因此可能会对 GCM 的不确定性造成放大。

虽然整体上 GCM 贡献的不确定性更大,但是整体上不确定性来源仍然是因地而异的。此外,不同指标之间也存在较大的差异,这主要是由于不同地方的作物生长状况与气候类型存在差异。例如 IOD 异常可能对欧亚大陆东部和西部有不同的影响,因此这两个地区的主要不确定性来源有较大差别。其次,不同的气候模型对物理过程有不同的表述,这可能导致它们模拟的未来气候存在差异 (McKenna and Maycock 2021)。正是由于这个物理过程的差异,也导致了大气环流场对局部气候的影响也有区别。所以,这也导致了产量与不同大气环流因子有不同的敏感性。对于作物模型来说,即使处于相同的气候条件下,作物模型的响应也可能存在差异。作物模型主导的地区可能是由于该地区大气环流因子与局部气候(生长季内)的相关性较弱所导致的。因此,虽然 GCM 仍然是主要的不确定性来源,但是不同地区和指标之间的差异也需要被充分考虑。

本研究首次考虑了这种关联分析的不确定性,但是并没有尝试约束不确定性。本研究建议,如果能在历史时期再现 ENSO 等因素对产量的影响,就可能改进预测的准确程度和约束不确定性。此外,模型选择和降尺度方法也可能对不确定性产生影响。因此,未来可能需要综合考虑多方面来降低不确定性。例如,对 ENSO 进行校准,可能有助于减少不确定性 (Tang et al. 2021)。

## 6.4 小结

本研究旨在识别和比较历史和未来全球作物产量变化的主要气候驱动因素,并可靠地估计全球和区域作物产量对气候驱动因素的敏感性,同时量化历史和未来大气环流因子强振荡阶段对产量的影响。该研究的不确定性及其主要来源也得到了探讨。研究结果表明:

(1) 在历史期间, ENSO 是小麦和玉米产量的主要驱动因素, 对欧洲和亚洲北部地区影响最大; 而 IOD 则是非洲和南美洲地区产量的主要驱动因素。未来期间(2000-2099), NAO 在北半球的玉米产量中占比例不断增大, 而在水稻和大豆中, NAO 同样有着明显的提升。

(2) 在 SSP585 情景下, IOD、ENSO 和 NAO 的正向异常导致产量有非常强烈的变化, 而负向异常变化不大, 尤其是 IOD 和 NAO。在 SSP126 阶段, 产量变化与历史相似, 但 ENSO 异常仍然导致产量有明显的变化。

(3) 对玉米而言, NAO 的负向异常和 ENSO 正向和负向的异常将会威胁到多个区域的产量; 对小麦而言, 应特别关注 IOD 异常; 对大豆和水稻而言, NAO 正向或负向的异常将对产量造成极大的减产。

(4) 总体而言, GCM 对不确定性的贡献高于 SSP, 而 SSP 对不确定性的贡献高于 GGCM。在水稻和玉米产量对 NAO 的敏感性分析中, SSP 的不确定性贡献高于 GCM。

本研究的发现有助于提高对气候变化对全球主要粮食产区的影响的理解, 有潜力提高全球粮食系统的韧性。



## 第七章 结论和展望

### 7.1 主要结论

本文基于不同方法在不同尺度上模拟和预测了产量变化以及气候变化对产量的影响。首先基于机器学习对产量进行预报和模拟，并找出不同变量对产量的影响。然后将机器学习与网格作物模型结合以提高模型精度并有效的约束了模型的不确定性。在这个基础上，本研究提出了一个新的框架来探讨产量预测以及不确定性，探讨了模型组合和模型数量对不确定性的影响。最后，建立了大尺度环流与产量的遥相关关系，并分析了气候变异对粮食安全的影响。本研究可以为决策者制定科学的管理措施提供有价值的参考，并且有助于提高对对气候变化与粮食安全的认识。本研究的主要结论如下：

#### (1) 基于机器学习的产量预测及其与环境变量的响应关系

通过建立小麦产量预测模型，本研究比较了不同机器学习方法和数据集的预测能力，发现随机森林模型优于支持向量机。采用基于植被指数  $NIR_v$  ( $R^2 = 0.74$ ;  $RMSE = 758$  kg/ha) 的随机森林模型比使用  $EVI$  ( $R^2 = 0.73$ ;  $RMSE = 762$  kg/ha) 或  $NDVI$  ( $R^2 = 0.73$ ;  $RMSE = 770$  kg/ha) 更为精确。本研究进一步探讨了各种环境变量对小麦产量的影响程度。结果显示，基于植被指数对小麦产量影响最大，其中 T3 阶段（拔节至抽穗）的  $NIR_v$  是最重要的预测因子。此外，辐射（Rad，阶段：分蘖至抽穗）、相对湿度（RHum，阶段：播种至分蘖）、土壤有机碳（SOC）、风速（Ws，阶段：整个生育阶段）、小雨（R5，阶段：播种至拔节）和霜冻（Frost，阶段：分蘖至拔节）被确定为限制小麦产量的主要因素。研究还发现，灌溉减轻了干旱对产量变化的贡献。本研究采用局部依赖曲线解释了每个生育阶段的不同预测因子如何影响小麦产量，发现  $NIR_v$  与小麦产量呈线性和非线性关系，而小麦产量对其他环境变量则呈现类似阈值的响应曲线。

#### (2) 建立产量预报模型并识别关键生长阶段

使用由多源环境变量驱动的机器学习可以提供精度较高的作物产量预测。在距离收获前约一到三个月的时间内，以下作物产量预测具有较好的模拟精度：冬小麦 ( $r=0.81-0.85$ ,  $nRMSE=10.5-11.4\%$ )、春玉米 ( $r=0.79-0.81$ ,  $nRMSE=17.1-17.9\%$ )、夏玉米 ( $r=0.77-0.79$ ,  $nRMSE=10.2-10.4\%$ )、早稻 ( $r=0.71-0.72$ ,  $nRMSE=7.4-7.5\%$ )、中稻 ( $r=0.78-0.82$ ,  $nRMSE=7.6-8.3\%$ ) 和晚稻 ( $r=0.76-0.78$ ,  $nRMSE=8.6-8.9\%$ )。本研究开发的机器学习预测模型在不同提前期为农民和决策者提供有价值的信息，有助于降低生长季节结束前产量损失的风险。此外，本研究还确定了影响小麦、玉米和水

稻产量的主要预测因子。总体而言，太阳辐射和植被指数（尤其是拔节至乳熟期间）被认定为冬小麦的主要预测因子；植被指数（整个生长季节）和干旱（特别是出苗至抽穗阶段）为春玉米的关键预测指标；土壤水分（整个生长季节）是夏玉米、晚稻和中稻的重要预测因子；而降水（特别是在孕穗期到抽穗期）是早稻的关键预报因子。

### （3）基于多模型集合预测气候变化对作物产量的影响及不确定性分析

气候模式驱动的作物模型已经被广泛应用于分析未来气候变化对农业生产的影响，然而在预测中存在较大的不确定性。目前，尚未明确研究模型组合和数量对这种不确定性的影响。本研究探讨了全球网格作物模型（GGCM）和全球气候模型（GCM）在模拟未来气候下，全球作物产量变化时，集合配置（即模型组成和数量）对建模不确定性的影响。研究表明，不同 GCM 和 GGCM 组合中的建模不确定性在气候变化背景下存在较大差异，尤其是对于玉米、水稻和大豆等作物，而小麦的不确定性相对较小。此外，随机选择大约 6 个 GGCM 和 10 个 GCM 可以基本反映 9 个作物模型和 32 个气候模型的整体不确定性。然而，基于聚类分析的结果表明，可以进一步减少最小集合数目。研究发现，在确保每个聚类至少包含一个模型的情况下，大约 3-4 个模型便可反映整体方差。这些发现突显了模型组合和集合规模在确定作物产量预测中不确定性的主要来源方面的重要性。

### （4）基于机器学习提高网格作物模型对产量的预测并约束不确定性

通过将机器学习与作物模型输出相结合进行改进，本研究可以显著提高模型的精度。校准前，玉米的  $R^2$  为 0.06-0.37，nRMSE 约为 0.2-0.4（仅 JULES 大于 50%）；大豆的  $R^2$  为 0.05-0.32，nRMSE 约为 0.2-0.35。然而，在校准后，玉米的  $R^2$  提高至 0.37-0.58，nRMSE 均小于 0.2；大豆的  $R^2$  提高至 0.22-0.37，nRMSE 波动在 0.2 左右。病虫害、高温和干旱是影响大豆产量的主要因素，而低温、病虫害和干旱则是玉米产量变化的主要影响因素。本研究的方法可以约束 48-67% 的不确定性。在 21 世纪中叶（T1 时段），作物模型是玉米产量不确定性的主要来源，而气候模式则是大豆产量不确定性的主要来源。到了 21 世纪末（T2 时段），排放情景成为主要的不确定性来源。值得注意的是，不确定性来源在不同区域存在一定的差异。

### （5）来情景下气候驱动因子对全球主要粮食作物产量的影响

对比单一气象要素，气候驱动因素所包含的信息更加丰富。因此，分析气候驱动因子与作物产量之间的关系能更准确地预测作物产量变化趋势，并深入理解气候变化对作物产量的影响机制。研究结果显示，在历史时期，ENSO 成为小麦和玉米产量的主要驱动因素，对欧洲和亚洲北部地区影响最为显著；而 IOD 则主导非洲和南美洲地区产量变化。在未来时期（2000-2099），NAO 在北半球玉米产量中的影响比例逐渐增加，并在水稻和大豆产量中也呈现明显上升趋势。在 SSP585 情景下，IOD、ENSO 和 NAO 的正向异常将导致产量剧烈波动，而负向异常的影响相对较小，特别是对于

IOD 和 NAO。在 SSP126 阶段，产量变化与历史相似，但 ENSO 异常仍可能引发明显产量波动。对玉米而言，NAO 的负向异常及 ENSO 正向和负向异常将对多区域产量构成威胁。小麦则需特别关注 IOD 异常。对大豆和水稻来说，NAO 的正向或负向异常都将导致产量显著减少。在不确定性方面，GCM 对产量变化的不确定性贡献大于 SSP，而 SSP 的不确定性贡献大于 GGCM。然而，在水稻和玉米产量对 NAO 敏感性分析中，SSP 的不确定性贡献高于 GCM。总之，本研究提高了对气候变化对全球主要粮食产区的影响的理解，并对提高全球粮食系统的韧性具有潜在的意义。

## 7.2 创新点

本研究的创新点主要包括：

(1) 通过建立基于机器学习的产量预测模型，本研究分析了不同环境变量对产量预测的潜力，并揭示了不同因子与作物产量的非线性响应关系。在此基础上，能够提前预测作物产量，并在作物收获期前 1-3 个月可靠地预估产量。通过分析哪个生长阶段的环境变量对模型精度更敏感，确定了最优的预报时段。同时，建议在这些关键生长阶段应加强农业管理，以提高产量。本研究结果能为农业生产提供实时的决策支持，帮助农民提高产量并规避风险。

(2) 为更好地挖掘多模型集合的信息，在本研究中构建了一个全新的框架。考虑了多模型集合中不同特性的模型组合以及模型数量对产量预测的影响，并量化了不同集合配置下的不确定性来源。此外，本研究首次采用机器学习与不同作物模型构建混合模型，以考虑病虫害和极端气候风险等不同环境变量对产量的影响，并提出了一个新的思路以约束产量预测的不确定性。

(3) 大气环流因子比单一气候变量往往包含更多信息。因此，为更好地揭示气候变化对产量的影响，本研究识别和比较了历史和未来全球不同区域影响作物产量的主要气候驱动因子，并量化了这些因子对产量的影响。同时，本研究可靠地估计了全球和区域作物产量对气候驱动因素的敏感性以及相关的不确定性来源。本研究结果为粮食生产、全球粮食贸易等提供了有价值的信息。

## 7.3 研究不足和展望

(1) 在研究中，统计模型是一种数据驱动的方法，因此数据的质量和数量对模型的性能起着至关重要的作用。如果数据质量不高或数量不足，那么模型的结果可能会受到影响，甚至可能出现错误的预测结果。这就要求我们在研究中尽可能采用高质量的数据，并确保数据的充足性。在农业生产中，田间管理方案是影响作物产量的关键因素之一，例如灌溉和施肥。然而，基于统计模型的预测种，这些往往被忽略掉了。

如果没有考虑和说明这些管理因素，那么对产量的预测就可能会出现偏差。因此，在研究中必须考虑这些因素，并尽可能将其纳入统计模型中以提高预测的准确性。此外，品种选择也是影响作物产量的重要因素。不同的基因型将对不同产区的产量产生显著影响，这需要在研究中加以考虑。如果我们没有考虑品种选择，那么可能会导致预测结果不准确，从而影响生产效益。因此，在农业生产中必须对品种选择进行充分的研究，以确保能够选择最适合当地环境和需求的品种。

(2) GGCM 模拟没有考虑土壤数据和管理选项（例如施肥率）等不确定性来源以及病虫害的影响。其次，模型参数化也会影响模拟并影响不确定性。虽然本研究中考虑了 GGCM 之间的结构差异，但没有考虑单个 GGCM 参数设置的不确定性。尽管一些 GGCM 可以捕捉极端温度事件（包括高温和霜冻）和干旱情况下发生的产量损失，但它们低估了极端潮湿条件下的产量损失程度。

(3) 鉴于目前的预测仍然有较大的不确定性，本研究以及当前一些其它研究中的都存在的一个不足是没有更好的去降低不确定性。虽然一些研究通过对模型改进以提高模型的模拟精度，例如改进温度响应(Wang et al. 2017b)，嵌入涝渍(Liu et al. 2023)或光合(Wang et al. 2022b)的影响模块。也有研究基于统计模型和实验数据对产量进行紧急约束，来有效的降低不确定性。因此，仍然需要开展控制变量实验，并将其有效的用于模型开发工作中。如何协调不同模型预测的结果也是一个关键的问题。虽然本研究提出一个框架可以通过建立不同子集来更充分考虑气候变化的影响。然而，更好的协调气候模型与过程模型来提高预测精度才能从根本上解决问题。后续将从不同角度，结合不同观测数据或数据集从不同角度来优化模型，也期待这从大量的模型预测中得到“确定”的结果。

(4) 目前已经有大量的工作分析了气候变化对产量的影响。例如极端降水(Lesk et al. 2020; Li et al. 2019c)，高温胁迫(Liu et al. 2016; Perry et al. 2020; Shew et al. 2020)，以及干旱等(Leng and Hall 2019; Li et al. 2022b)。还有一些研究考虑了一些复合事件：湿度-温度(Lesk et al. 2021)，高温-干旱(Heino et al. 2022)，高温-极端湿润(Lesk et al. 2022)，大气-土壤复合干旱等(Zhou et al. 2019a; Zhou et al. 2019b)。而这些极端事件的形成往往受一些大气环流的影响。本研究虽然分析了 ENSO、IOD 和 NAO 历史和未来与产量的关系。然而，由于环流指数对产量影响的转变可能也会导致粮食贸易格局在不同年份发生变化（例如厄尔尼诺年与拉尼娜年可能不同）。因此，下一步研究将结合一些经济模型，分析环流指数异常对粮食生产相关的经济政策方面的影响。以帮助决策者和利益相关者识别不同政策的潜在风险、机遇和权衡，并就如何改善粮食安全和可持续性做出合适的决定。这样可以为资源配置、投资和贸易政策提供指导，以减轻气候变化导致的负面影响。

## 参考文献

- 车少静, 智利辉 和 冯立辉, 2005. 气候变暖对石家庄冬小麦主要生育期的影响及对策. 中国农业气象, 26(3): 180-183.
- 陈浩, 许月萍, 郑超昊 和 郭玉雪, 2023. 厄尔尼诺-南方涛动影响下月尺度水文模型参数可移植性. 同济大学学报:自然科学版, 51(1): 8.
- 段安民 和 张萍, 2022. 青藏高原大气热源年际变率及其驱动因子. 大气科学学报, 46(2): 18.
- 郭品文 和 吴清传, 2022. 印度洋偶极子向海盆一致模转变的年代际变化及其成因. 大气科学学报, 45(4): 11.
- 靳彦华, 熊黑钢, 张芳 和 王莉峰, 2013. 不同地类春小麦叶片叶绿素含量高光谱植被指数估算模型研究. 麦类作物学报.
- 竞霞, 邹琴, 白宗璠 和 黄文江, 2021. 基于反射光谱和叶绿素荧光数据的作物病害遥感监测研究进展. 作物学报, 47(11): 2067-2079.
- 刘苇航, 叶涛, 史培军 和 陈说, 2022. 气候变化对粮食生产风险的影响研究进展. 自然灾害学报(004): 031.
- 刘兴华, 季致建, 吴洪宝 和 于秀晶, 2006. 中国近 40 年极端气温和降水的分布特征及年代际差异. 热带气象学报, 22(6): 618-624.
- 莫嘉琪, 王辉 和 林万涛, 2006. 厄尔尼诺-南方涛动时滞海-气振子耦合模型. 物理学报, 55(7): 4.
- 彭慧文, 赵俊芳, 谢鸿飞 和 房世波, 2022. 作物模型应用与遥感信息集成技术研究进展. 中国农业气象, 43(8): 13.
- 宋承国, 2009. 世界粮食危机与中国粮食安全. 当代经济研究(2): 4.
- 隋学艳, 薛光山, 王素娟, 侯学会 和 张晓冬, 2016. 基于 MODIS 数据的山东省农作物旱情遥感监测报告. 山东农业科学, 48(002): 129-132.
- 孙晶晶, 赵凯, 曹慧 和 牛影影, 2018. 我国耕地保护经济补偿分区及其补偿额度测算——基于省级耕地-经济协调性视角. 浙江国土资源.
- 孙擎, 杨再强, 车向红, 杨世琼, 王琳 和 郑晓辉, 2019. 水稻格点作物模型在中国区域的不确定性评估. 中国农业气象, 40(3): 14.
- 王劲松, 李耀辉, 王润元, 冯建英 和 赵艳霞, 2012. 我国气象干旱研究进展评述. 干旱气象, 30(4): 497-508.
- 吴普特 和 赵西宁, 2010. 气候变化对中国农业用水和粮食生产的影响. 中国农业工程学报, 26(2): 1-6.
- 薛金儒 和 吕肖良, 2022. 黄土高原生态工程实施下基于日光诱导叶绿素荧光的植被恢复生产力效益评价. 植物生态学报, 46(10): 1289-1304.
- 严海军, 卓越, 李茂娜, 王云玲, 郭辉, 王晶晶, 李长硕 和 丁峰, 2022. 基于机器学习和无人机多光谱遥感的苜蓿产量预测. 农业工程学报, 38(11): 8.
- 杨伟才 和 毛晓敏, 2018. 气候变化影响下作物模型的不确定性. 排灌机械工程学报, 36(9): 7.
- 翟石艳, 秦耀辰 和 宋根鑫, 2017. 气候变化背景下农业适应性研究进展. 河南大学学报(自然版)(2): 136.

- 张朝, 张亮亮 和 陶福祿, 2020. 耦合作物模型和机器学习的作物育种适应时间预测方法.
- 张菊, 房世波 和 刘汉湖, 2021. 基于微波数据与光学数据集成的机器学习技术在作物产量估算中的应用. *地球信息科学学报*, 23(6): 1082-1091.
- 张宁宁, 李雪, 吕新业 和 袁龙江, 2022. 百年变局,世纪疫情背景下世界及中国粮食安全面临的风险挑战及应对策略. *农业经济问题*(12): 6.
- 赵松乔, 1983. 中国综合自然地理区划的一个新方案. *地理学报*, 50(1).
- 钟仁海, 2022. 基于深度学习的玉米单产估测方法和归因分析研究. 浙江大学.
- 周营, 2018. 我国耕地保护面临的严峻形势和政策性建议. *西部资源*, 000(3): 190-191.
- 朱晶, 张瑞华 和 谢超平, 2022. 全球农业贸易治理与中国粮食安全. *农业经济问题*, 43(11): 14.
- Alexandrov, V.A. and Hoogenboom, G., 2000. The impact of climate variability and change on crop yield in Bulgaria. *Agricultural and Forest Meteorology*, 104(4): 315-327.
- Anderson, W., Baethgen, W., Capitanio, F., Ciais, P., Cook, B., da Cunha, G.R., Goddard, L., Schauburger, B., Sonder, K. and Van der Velde, M., 2021. The past, present, and future of multiple wheat and maize breadbasket shocks.
- Anderson, W., Baethgen, W., Capitanio, F., Ciais, P., Cook, B.I., Cunha, C.G.R.d., Goddard, L., Schauburger, B., Sonder, K., Podestá, G., van der Velde, M. and You, L., 2023. Climate variability and simultaneous breadbasket yield shocks as observed in long-term yield records. *Agricultural and Forest Meteorology*, 331.
- Anderson, W., Seager, R., Baethgen, W., Cane, M. and You, L., 2019. Synchronous crop failures and climate-forced production variability. *Science advances*, 5(7): eaaw1976.
- Anderson, W.B., Han, E., Baethgen, W., Goddard, L., Muñoz, Á.G. and Robertson, A.W., 2020. The Madden - Julian Oscillation Affects Maize Yields Throughout the Tropics and Subtropics. *Geophysical Research Letters*, 47(11).
- Anna, C., Salah, S. and Brett, W., 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers Electronics in Agriculture*, 151: 61-69.
- Anttila-Hughes, J.K., Jina, A.S. and McCord, G.C., 2021. ENSO impacts child undernutrition in the global tropics. *Nat Commun*, 12(1): 5785.
- Asch, F., Dingkuhn, M., Wittstock, C. and Doerffling, K., 1999. Sodium and potassium uptake of rice panicles as affected by salinity and season in relation to yield and yield components. *Plant and Soil*, 207(2): 133-145.
- Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V.V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurrealde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z. and Zhu, Y., 2014. Rising temperatures reduce global wheat production. *Nature Climate Change*, 5(2): 143-147.
- Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J.,



- Rötter, R.P. and Cammarano, D., 2013. Uncertainty in simulating wheat yields under climate change. *Nature climate change*, 3(9): 827-832.
- Asseng, S., Guarin, J.R., Raman, M., Monje, O., Kiss, G., Despommier, D.D., Meggers, F.M. and Gauthier, P.P.G., 2020. Wheat yield potential in controlled-environment vertical farms. *Proc Natl Acad Sci U S A*, 117(32): 19131-19135.
- Atteya, A., 2003. Alteration of water relations and yield of corn genotypes in response to drought stress. *Bulg. J. Plant Physiol*, 29(1-2): 63-76.
- Bacour, C., Maignan, F., Peylin, P., MacBean, N., Bastrikov, V., Joiner, J., Köhler, P., Guanter, L. and Frankenberg, C., 2019. Differences Between OCO - 2 and GOME - 2 SIF Products From a Model - Data Fusion Perspective. *Journal of Geophysical Research: Biogeosciences*, 124(10): 3143-3157.
- Badgley, G., Anderegg, L.D.L., Berry, J.A. and Field, C.B., 2019. Terrestrial gross primary production: Using NIRV to scale from site to globe. *Glob Chang Biol*, 25(11): 3731-3740.
- Badgley, G., Field, C.B. and Berry, J.A.J.S.A., 2017. Canopy near-infrared reflectance and terrestrial photosynthesis. 3(3): e1602244.
- Barlow, K.M., Christy, B.P., O'Leary, G.J., Riffkin, P.A. and Nuttall, J.G., 2015. Simulating the impact of extreme heat and frost events on wheat crop production: A review. *Field Crops Research*, 171: 109-119.
- Bastos, A., Janssens, I.A., Gouveia, C.M., Trigo, R.M., Ciais, P., Chevallier, F., Penuelas, J., Rodenbeck, C., Piao, S., Friedlingstein, P. and Running, S.W., 2016. European land CO<sub>2</sub> sink influenced by NAO and East-Atlantic Pattern coupling. *Nat Commun*, 7: 10315.
- Bastos, L.M., Carciocchi, W., Lollato, R.P., Jaenisch, B.R., Rezende, C.R., Schwalbert, R., Vara Prasad, P.V., Zhang, G., Fritz, A.K., Foster, C., Wright, Y., Young, S., Bradley, P. and Ciampitti, I.A., 2020. Winter Wheat Yield Response to Plant Density as a Function of Yield Environment and Tillering Potential: A Review and Field Studies. *Front Plant Sci*, 11: 54.
- Becker, R., Chambers, J. and Wilks, A., 1988. *The New S Language*. Wadsworth & Brooks/Cole. Computer Science Series, Pacific Grove, CA.
- Bergjord, A.K., Bonesmo, H. and Skjelvåg, A.O., 2008. Modelling the course of frost tolerance in winter wheat. *European Journal of Agronomy*, 28(3): 321-330.
- Besalatpour, A., Ayoubi, S., Hajabbasi, M., Jazi, A.Y. and Gharipour, A., 2014. Feature selection using parallel genetic algorithm for the prediction of geometric mean diameter of soil aggregates by machine learning methods. *Arid Land Research Management*, 28(4): 383-394.
- Besalatpour, A., Hajabbasi, M., Ayoubi, S., Gharipour, A. and Jazi, A.Y., 2012. Prediction of soil physical properties by optimized support vector machines. *International Agrophysics*, 26(2).
- Blum, A., Ramaiah, S., Kanemasu, E. and Paulsen, G., 1990. Wheat recovery from drought stress at the tillering stage of development. *Field Crops Research*, 24(1-2): 67-85.
- Bonnet, R., Swingedouw, D., Gastineau, G., Boucher, O., Deshayes, J., Hourdin, F., Mignot, J., Servonnat, J. and Sima, A., 2021. Increased risk of near term global warming due to a recent AMOC weakening. *Nat Commun*, 12(1): 6108.
- Borg, I. and Groenen, P.J., 2005. *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media.

- Breiman, L., 2001. Random forests. *Machine learning*, 45(1): 5-32.
- Brown, J.N., Hochman, Z., Holzworth, D. and Horan, H., 2018. Seasonal climate forecasts provide more definitive and accurate crop yield predictions. *Agricultural and Forest Meteorology*, 260-261: 247-254.
- Burgess, A.J., Gibbs, J.A. and Murchie, E.H., 2019. A canopy conundrum: can wind-induced movement help to increase crop productivity by relieving photosynthetic limitations? *J Exp Bot*, 70(9): 2371-2380.
- Cai, W., Ng, B., Wang, G., Santoso, A., Wu, L. and Yang, K., 2022. Increased ENSO sea surface temperature variability under four IPCC emission scenarios. *Nature Climate Change*, 12(3): 228-231.
- Cai, W., Santoso, A., Wang, G., Wu, L., Collins, M., Lengaigne, M., Power, S. and Timmermann, A., 2020. ENSO response to greenhouse forcing. *El Niño Southern Oscillation in a Changing Climate*: 289-307.
- Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L. and Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agricultural and Forest Meteorology*, 274: 144-159.
- Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M.D., Muñoz-Marí, J., García-Haro, F.J. and Guanter, L., 2021. A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances*, 7(9): eabc7447.
- Cao, J., Zhang, Z., Luo, Y., Zhang, L., Zhang, J., Li, Z. and Tao, F., 2021b. Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. *European Journal of Agronomy*, 123.
- Cao, J., Zhang, Z., Tao, F., Chen, Y., Luo, X. and Xie, J., 2023. Forecasting global crop yields based on El Niño Southern Oscillation early signals. *Agricultural Systems*, 205.
- Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Han, J. and Li, Z., 2020. Identifying the Contributions of Multi-Source Data for Winter Wheat Yield Prediction in China. *Remote Sensing*, 12(5).
- Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Zhang, J., Han, J. and Xie, J., 2021a. Integrating Multi-Source Data for Rice Yield Prediction across China using Machine Learning and Deep Learning Approaches. *Agricultural and Forest Meteorology*, 297.
- Chakrabarti, S., Bongiovanni, T., Judge, J., Zotarelli, L. and Bayer, C., 2014. Assimilation of SMOS Soil Moisture for Quantifying Drought Impacts on Crop Yield in Agricultural Regions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(9): 3867-3879.
- Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R. and Chhetri, N., 2014. A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, 4(4): 287-291.
- Chang, K.-W., Shen, Y. and Lo, J.-C., 2005. Predicting Rice Yield Using Canopy Reflectance Measured at Booting Stage. *Agronomy Journal*, 97(3): 872-878.
- Chapagain, R., Remenyi, T.A., Harris, R.M.B., Mohammed, C.L., Huth, N., Wallach, D., Rezaei, E.E. and Ojeda, J.J., 2022. Decomposing crop model uncertainty: A systematic review. *Field Crops Research*, 279.
- Chaves, M.M., Pereira, J.S., Maroco, J., Rodrigues, M.L., Ricardo, C.P., Osorio, M.L., Carvalho, I., Faria, T. and Pinheiro, C., 2002. How plants cope with water stress in the field. Photosynthesis and growth. *Ann Bot*, 89 Spec No: 907-16.
- Chen, A., Mao, J., Ricciuto, D., Xiao, J., Frankenberg, C., Li, X., Thornton, P.E., Gu, L. and Knapp, A.K.,

- 2020a. Moisture availability mediates the relationship between terrestrial gross primary production and solar-induced chlorophyll fluorescence: Insights from global-scale variations. *Glob Chang Biol*.
- Chen, C., Baethgen, W.E. and Robertson, A., 2012. Contributions of individual variation in temperature, solar radiation and precipitation to crop yield in the North China Plain, 1961–2003. *Climatic Change*, 116(3-4): 767-788.
- Chen, H., Li, L., Luo, X., Li, Y., Liu, D.L., Zhao, Y., Feng, H. and Deng, J., 2019. Modeling impacts of mulching and climate change on crop production and N<sub>2</sub>O emission in the Loess Plateau of China. *Agricultural and Forest Meteorology*, 268: 86-97.
- Chen, S., Jiang, T., Ma, H., He, C., Xu, F., Malone, R.W., Feng, H., Yu, Q., Siddique, K.H.M., Dong, Q.g. and He, J., 2020b. Dynamic within-season irrigation scheduling for maize production in Northwest China: A Method Based on Weather Data Fusion and yield prediction by DSSAT. *Agricultural and Forest Meteorology*, 285-286.
- Chen, X., Li, Y., Yao, N., Liu, D.L., Javed, T., Liu, C. and Liu, F., 2020d. Impacts of multi-timescale SPEI and SMDI variations on winter wheat yields. *Agricultural Systems*, 185.
- Chen, X., Wang, L., Niu, Z., Zhang, M., Li, C.a. and Li, J., 2020c. The effects of projected climate change and extreme climate on maize and rice in the Yangtze River Basin, China. *Agricultural and Forest Meteorology*, 282-283.
- Chen, Y., Zhang, Z. and Tao, F., 2018. Improving regional winter wheat yield estimation through assimilation of phenology and leaf area index from remote sensing data. *European journal of agronomy*, 101: 163-173.
- Collins, M., An, S.-I., Cai, W., Ganachaud, A., Guilyardi, E., Jin, F.-F., Jochum, M., Lengaigne, M., Power, S., Timmermann, A., Vecchi, G. and Wittenberg, A., 2010. The impact of global warming on the tropical Pacific Ocean and El Niño. *Nature Geoscience*, 3(6): 391-397.
- Cucchi, M., Weedon, G.P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H. and Buontempo, C., 2020. WFDE5: bias-adjusted ERA5 reanalysis data for impact studies. *Earth System Science Data*, 12(3): 2097-2120.
- Dawson, I. and Wardlaw, I., 1989. The tolerance of wheat to high temperatures during reproductive growth. III. Booting to anthesis. *Australian Journal of Agricultural Research*, 40(5): 965-980.
- Dechant, B., Ryu, Y., Badgley, G., Köhler, P., Rascher, U., Migliavacca, M., Zhang, Y., Tagliabue, G., Guan, K. and Rossini, M., 2020b. NIRvP: a robust structural proxy for sun-induced chlorophyll fluorescence and photosynthesis across scales.
- Dechant, B., Ryu, Y., Badgley, G., Zeng, Y., Berry, J.A., Zhang, Y., Goulas, Y., Li, Z., Zhang, Q., Kang, M., Li, J. and Moya, I., 2020a. Canopy structure explains the relationship between photosynthesis and sun-induced chlorophyll fluorescence in crops. *Remote Sensing of Environment*, 241.
- Dokoohaki, H., Kivi, M.S., Martinez-Feria, R., Miguez, F.E. and Hoogenboom, G., 2021. A comprehensive uncertainty quantification of large-scale process-based crop modeling frameworks. *Environmental Research Letters*, 16(8).
- Duveiller, G., Filippini, F., Walther, S., Köhler, P., Frankenberg, C., Guanter, L. and Cescatti, A., 2020. A spatially downscaled sun-induced fluorescence global product for enhanced monitoring of vegetation productivity. *Earth System Science Data*, 12(2): 1101-1116.

- Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K.J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R.C., Mueller, N.D., Ray, D.K., Rosenzweig, C., Ruane, A.C. and Sheffield, J., 2015. The Global Gridded Crop Model Intercomparison: data and modeling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8(2): 261-277.
- Fageria, N.K., 2007. Yield Physiology of Rice. *Journal of Plant Nutrition*, 30(6): 843-879.
- Falconnier, G.N., Corbeels, M., Boote, K.J., Affholder, F., Adam, M., MacCarthy, D.S., Ruane, A.C., Nendel, C., Whitbread, A.M., Justes, E., Ahuja, L.R., Akinseye, F.M., Alou, I.N., Amouzou, K.A., Anapalli, S.S., Baron, C., Basso, B., Baudron, F., Bertuzzi, P., Challinor, A.J., Chen, Y., Deryng, D., Elsayed, M.L., Faye, B., Gaiser, T., Galdos, M., Gayler, S., Gerardeaux, E., Giner, M., Grant, B., Hoogenboom, G., Ibrahim, E.S., Kamali, B., Kersebaum, K.C., Kim, S.H., van der Laan, M., Leroux, L., Lizaso, J.I., Maestrini, B., Meier, E.A., Mequanint, F., Ndoli, A., Porter, C.H., Priesack, E., Ripoche, D., Sida, T.S., Singh, U., Smith, W.N., Srivastava, A., Sinha, S., Tao, F., Thorburn, P.J., Timlin, D., Traore, B., Twine, T. and Webber, H., 2020. Modelling climate change impacts on maize yields under low nitrogen input conditions in sub-Saharan Africa. *Glob Chang Biol*, 26(10): 5942-5964.
- FAO, 2018. *World Food and Agriculture—Statistical Pocketbook*.
- Feng, B., Zhuo, Mekonnen, M.M., Marston, L.T., Yang, X., Xu, Z., Liu, Y., Wang, W., Li, Z., Li, M., Ji, X. and Wu, P., 2022a. Inputs for staple crop production in China drive burden shifting of water and carbon footprints transgressing part of provincial planetary boundaries. *Water Res*, 221: 118803.
- Feng, P., Wang, B., Liu, D.L., Waters, C. and Yu, Q., 2019. Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia. *Agricultural and Forest Meteorology*, 275: 100-113.
- Feng, P., Wang, B., Liu, D.L., Waters, C., Xiao, D., Shi, L. and Yu, Q., 2020. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. *Agricultural and Forest Meteorology*, 285-286.
- Feng, P., Wang, B., Macadam, I., Taschetto, A.S., Abram, N.J., Luo, J.-J., King, A.D., Chen, Y., Li, Y., Liu, D.L., Yu, Q. and Hu, K., 2022b. Increasing dominance of Indian Ocean variability impacts Australian wheat yields. *Nature Food*.
- Fereday, D.R., Chadwick, R., Knight, J.R. and Scaife, A.A., 2020. Tropical Rainfall Linked to Stronger Future ENSO - NAO Teleconnection in CMIP5 Models. *Geophysical Research Letters*, 47(22).
- Folberth, C., Elliott, J., Muller, C., Balkovic, J., Chryssanthacopoulos, J., Izaurralde, R.C., Jones, C.D., Khabarov, N., Liu, W., Reddy, A., Schmid, E., Skalsky, R., Yang, H., Arneth, A., Ciais, P., Deryng, D., Lawrence, P.J., Olin, S., Pugh, T.A.M., Ruane, A.C. and Wang, X., 2019. Parameterization-induced uncertainties and impacts of crop management harmonization in a global gridded crop model ensemble. *PLoS One*, 14(9): e0221862.
- Folberth, C., Skalsky, R., Moltchanova, E., Balkovic, J., Azevedo, L.B., Obersteiner, M. and van der Velde, M., 2016. Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. *Nat Commun*, 7: 11872.
- Franke, J.A., Müller, C., Elliott, J., Ruane, A.C., Jägermeyr, J., Snyder, A., Dury, M., Falloon, P.D., Folberth, C., François, L., Hank, T., Izaurralde, R.C., Jacquemin, I., Jones, C., Li, M., Liu, W., Olin, S., Phillips, M., Pugh, T.A.M., Reddy, A., Williams, K., Wang, Z., Zabel, F. and Moyer, E.J., 2020a. The GGCM

- Phase-2 emulators: global gridded crop model responses to changes in CO<sub>2</sub>, temperature, water, and nitrogen (version 1.0). *Geoscientific Model Development*, 13(9): 3995-4018.
- Franke, J.A., Müller, C., Elliott, J., Ruane, A.C., Jägermeyr, J., Balkovic, J., Ciais, P., Dury, M., Falloon, P.D., Folberth, C., François, L., Hank, T., Hoffmann, M., Izaurrealde, R.C., Jacquemin, I., Jones, C., Khabarov, N., Koch, M., Li, M., Liu, W., Olin, S., Phillips, M., Pugh, T.A.M., Reddy, A., Wang, X., Williams, K., Zabel, F. and Moyer, E.J., 2020b. The GGCM Phase 2 experiment: global gridded crop model simulations under uniform changes in CO<sub>2</sub>, temperature, water, and nitrogen levels (protocol version 1.0). *Geoscientific Model Development*, 13(5): 2315-2336.
- Franke, J.A., Müller, C., Minoli, S., Elliott, J., Folberth, C., Gardner, C., Hank, T., Izaurrealde, R.C., Jägermeyr, J., Jones, C.D., Liu, W., Olin, S., Pugh, T.A.M., Ruane, A.C., Stephens, H., Zabel, F. and Moyer, E.J., 2022. Agricultural breadbaskets shift poleward given adaptive farmer behavior under climate change. *Glob Chang Biol*, 28(1): 167-181.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*: 1189-1232.
- Galik, C.S., 2019. A continuing need to revisit BECCS and its potential. *Nature Climate Change*, 10(1): 2-3.
- Gao, Y., Zhao, H., Zhao, C., Hu, G., Zhang, H., Liu, X., Li, N., Hou, H. and Li, X., 2022. Spatial and temporal variations of maize and wheat yield gaps and their relationships with climate in China. *Agricultural Water Management*, 270.
- Gao, Z., Liang, X.-G., Lin, S., Zhao, X., Zhang, L., Zhou, L.-L., Shen, S. and Zhou, S.-L., 2017. Supplemental irrigation at tasseling optimizes water and nitrogen distribution for high-yield production in spring maize. *Field Crops Research*, 209: 120-128.
- Gaupp, F., Hall, J., Hochrainer-Stigler, S. and Dadson, S., 2019. Changing risks of simultaneous global breadbasket failure. *Nature Climate Change*, 10(1): 54-57.
- Gohar, A.A. and Cashman, A., 2016. A methodology to assess the impact of climate variability and change on water resources, food security and economic welfare. *Agricultural Systems*, 147: 51-64.
- Gray, S.B., Dermody, O., Klein, S.P., Locke, A.M., McGrath, J.M., Paul, R.E., Rosenthal, D.M., Ruiz-Vera, U.M., Siebers, M.H., Strellner, R., Ainsworth, E.A., Bernacchi, C.J., Long, S.P., Ort, D.R. and Leakey, A.D., 2016. Intensifying drought eliminates the expected benefits of elevated carbon dioxide for soybean. *Nat Plants*, 2(9): 16132.
- Guo, Y., Fu, Y., Hao, F., Zhang, X., Wu, W., Jin, X. and Bryant, C.R., 2020. Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecological Indicators*, 120: 106935.
- Guo, Y., Wang, R., Tong, Z., Liu, X. and Zhang, J., 2019. Dynamic Evaluation and Regionalization of Maize Drought Vulnerability in the Midwest of Jilin Province. *Sustainability*, 11(15).
- Hammad, H.M., Khaliq, A., Abbas, F., Farhad, W., Fahad, S., Aslam, M., Shah, G.M., Nasim, W., Mubeen, M. and Bakhat, H.F., 2020. Comparative effects of organic and inorganic fertilizers on soil organic carbon and wheat productivity under arid region. *Communications in Soil Science Plant Analysis*, 51(10): 1406-1422.
- Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z. and Zhang, J., 2020. Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. *Remote Sensing*, 12(2).
- Harvell, C.D., Mitchell, C.E., Ward, J.R., Altizer, S., Dobson, A.P., Ostfeld, R.S. and Samuel, M.D., 2002.

- Climate warming and disease risks for terrestrial and marine biota. *Science*, 296(5576): 2158-2162.
- Hasegawa, T., Sakurai, G., Fujimori, S., Takahashi, K., Hijioka, Y. and Masui, T., 2021. Extreme climate events increase risk of global food insecurity and adaptation needs. *Nature Food*, 2(8): 587-595.
- He, L., Magney, T., Dutta, D., Yin, Y., Köhler, P., Grossmann, K., Stutz, J., Dold, C., Hatfield, J., Guan, K., Peng, B. and Frankenberg, C., 2020. From the Ground to Space: Using Solar - Induced Chlorophyll Fluorescence to Estimate Crop Productivity. *Geophysical Research Letters*, 47(7).
- He, Q., Liu, D.L., Wang, B., Li, L., Cowie, A., Simmons, A., Zhou, H., Tian, Q., Li, S., Li, Y., Liu, K., Yan, H., Harrison, M.T., Feng, P., Waters, C., Li, G.D., de Voil, P. and Yu, Q., 2022. Identifying effective agricultural management practices for climate change adaptation and mitigation: A win-win strategy in South-Eastern Australia. *Agricultural Systems*, 203.
- Hein, A.M., Gil, M.A., Twomey, C.R., Couzin, I.D. and Levin, S.A., 2018. Conserved behavioral circuits govern high-speed decision-making in wild fish shoals. *Proc Natl Acad Sci U S A*, 115(48): 12224-12228.
- Heinicke, S., Frieler, K., Jägermeyr, J. and Mengel, M., 2022. Global gridded crop models underestimate yield responses to droughts and heatwaves. *Environmental Research Letters*, 17(4).
- Heino, M., Guillaume, J.H.A., Müller, C., Iizumi, T. and Kummu, M., 2020. A multi-model analysis of teleconnected crop yield variability in a range of cropping systems. *Earth System Dynamics*, 11(1): 113-128.
- Heino, M., Kinnunen, P., Anderson, W., Ray, D., Puma, M.J., Varis, O., Siebert, S. and Kummu, M., 2022. Increased probability of hot and dry weather extremes during the growing season threatens global crop yields.
- Heino, M., Kinnunen, P., Anderson, W., Ray, D.K., Puma, M.J., Varis, O., Siebert, S. and Kummu, M., 2023. Increased probability of hot and dry weather extremes during the growing season threatens global crop yields. *Sci Rep*, 13(1): 3583.
- Heino, M., Puma, M.J., Ward, P.J., Gerten, D., Heck, V., Siebert, S. and Kummu, M., 2018. Two-thirds of global cropland area impacted by climate oscillations. *Nat Commun*, 9(1): 1257.
- Hoffman, A., R Kemanian, A. and E Forest, C., 2020. The response of maize, sorghum, and soybean yield to growing-phase climate revealed with machine learning. *Environmental Research Letters*, 15(9).
- Hori, M.E. and Ueda, H., 2006. Impact of global warming on the East Asian winter monsoon as revealed by nine coupled atmosphere-ocean GCMs. *Geophysical Research Letters*, 33(3).
- Hsiang, S.M., Meng, K.C. and Cane, M.A., 2011. Civil conflicts are associated with the global climate. *Nature*, 476(7361): 438-41.
- Huang, M., Piao, S., Ciais, P., Penuelas, J., Wang, X., Keenan, T.F., Peng, S., Berry, J.A., Wang, K., Mao, J., Alkama, R., Cescatti, A., Cuntz, M., De Deurwaerder, H., Gao, M., He, Y., Liu, Y., Luo, Y., Myneni, R.B., Niu, S., Shi, X., Yuan, W., Verbeeck, H., Wang, T., Wu, J. and Janssens, I.A., 2019. Air temperature optima of vegetation productivity across global biomes. *Nat Ecol Evol*, 3(5): 772-779.
- Huang, M., Wang, J., Wang, B., Liu, D.L., Feng, P., Yu, Q., Pan, X., Li, S. and Jiang, T., 2022. Dominant sources of uncertainty in simulating maize adaptation under future climate scenarios in China. *Agricultural Systems*, 199.
- Huang, M., Wang, J., Wang, B., Liu, D.L., Yu, Q., He, D., Wang, N. and Pan, X., 2020. Optimizing sowing



- window and cultivar choice can boost China's maize yield under 1.5 °C and 2 °C global warming. *Environmental Research Letters*, 15(2).
- Iizumi, T., Luo, J.J., Challinor, A.J., Sakurai, G., Yokozawa, M., Sakuma, H., Brown, M.E. and Yamagata, T., 2014. Impacts of El Niño Southern Oscillation on the global yields of major crops. *Nat Commun*, 5: 3712.
- Iizumi, T., Sakuma, H., Yokozawa, M., Luo, J.-J., Challinor, A.J., Brown, M.E., Sakurai, G. and Yamagata, T., 2013. Prediction of seasonal climate-induced variations in global food production. *Nature Climate Change*, 3(10): 904-908.
- Islam, F.S., Gault, A.G., Boothman, C., Polya, D.A., Charnock, J.M., Chatterjee, D. and Lloyd, J.R., 2004. Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. *Nature*, 430(6995): 68-71.
- Iturbide, M., Gutiérrez, J.M., Alves, L.M., Bedia, J., Cerezo-Mota, R., Gimenez, E., Cofiño, A.S., Di Luca, A., Faria, S.H., Gorodetskaya, I.V., Hauser, M., Herrera, S., Hennessy, K., Hewitt, H.T., Jones, R.G., Krakovska, S., Manzanar, R., Martínez-Castro, D., Narisma, G.T., Nurhati, I.S., Pinto, I., Seneviratne, S.I., van den Hurk, B. and Vera, C.S., 2020. An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets. *Earth System Science Data*, 12(4): 2959-2970.
- Jägermeyr, J., Müller, C., Ruane, A.C., Elliott, J., Balkovic, J., Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J.A., Fuchs, K., Guarin, J.R., Heinke, J., Hoogenboom, G., Iizumi, T., Jain, A.K., Kelly, D., Khabarov, N., Lange, S., Lin, T.-S., Liu, W., Mialyk, O., Minoli, S., Moyer, E.J., Okada, M., Phillips, M., Porter, C., Rabin, S.S., Scheer, C., Schneider, J.M., Schyns, J.F., Skalsky, R., Smerald, A., Stella, T., Stephens, H., Webber, H., Zabel, F. and Rosenzweig, C., 2021. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nature Food*.
- Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.-M., Gerber, J.S. and Reddy, V.R., 2016. Random forests for global and regional crop yield predictions. *PLoS One*, 11(6): e0156571.
- Jha, P.K., Athanasiadis, P., Gualdi, S., Trabucco, A., Mereu, V., Shelia, V. and Hoogenboom, G., 2019. Using daily data from seasonal forecasts in dynamic crop models for yield prediction: A case study for rice in Nepal's Terai. *Agricultural and Forest Meteorology*, 265: 349-358.
- Jiang, H., Hu, H., Zhong, R., Xu, J., Xu, J., Huang, J., Wang, S., Ying, Y. and Lin, T., 2020. A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level. *Global Change Biology*, 26(3): 1754-1766.
- Jiang, J. and Zhou, T., 2023. Agricultural drought over water-scarce Central Asia aggravated by internal climate variability. *Nature Geoscience*.
- Jiang, T., Wang, B., Xu, X., Cao, Y., Liu, D.L., He, L., Jin, N., Ma, H., Chen, S., Zhao, K., Feng, H., Yu, Q., He, Y. and He, J., 2022. Identifying sources of uncertainty in wheat production projections with consideration of crop climatic suitability under future climate. *Agricultural and Forest Meteorology*, 319.
- Jones, P.D., Jónsson, T. and Wheeler, D., 1997. Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and south - west Iceland. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 17(13): 1433-1450.

- Kamali, B., Lorite, I.J., Webber, H.A., Rezaei, E.E., Gabaldon-Leal, C., Nendel, C., Siebert, S., Ramirez-Cuesta, J.M., Ewert, F. and Ojeda, J.J., 2022. Uncertainty in climate change impact studies for irrigated maize cropping systems in southern Spain. *Sci Rep*, 12(1): 4049.
- Kersebaum, K.C., 2022. Modelling to Evaluate Climate Resilience of Crop Rotations Under Climate Change, *Climate Adaptation Modelling*. Springer, pp. 87-93.
- Kimball, B.A., 2016. Crop responses to elevated CO<sub>2</sub> and interactions with H<sub>2</sub>O, N, and temperature. *Current opinion in plant biology*, 31: 36-43.
- King, A.D., Pitman, A.J., Henley, B.J., Ukkola, A.M. and Brown, J.R., 2020. The role of climate variability in Australian drought. *Nature Climate Change*, 10(3): 177-179.
- Klueter, A., Crandall, J.B., Archer, F.I., Teece, M.A. and Coffroth, M.A., 2015. Taxonomic and environmental variation of metabolite profiles in marine dinoflagellates of the genus symbiodinium. *Metabolites*, 5(1): 74-99.
- Knutti, R., 2010. The end of model democracy? *Climatic Change*, 102(3-4): 395-404.
- Kothari, K., Battisti, R., Boote, K.J., Archontoulis, S.V., Confalone, A., Constantin, J., Cuadra, S.V., Debaeke, P., Faye, B., Grant, B., Hoogenboom, G., Jing, Q., van der Laan, M., Macena da Silva, F.A., Marin, F.R., Nehbandani, A., Nendel, C., Purcell, L.C., Qian, B., Ruane, A.C., Schoving, C., Silva, E.H.F.M., Smith, W., Soltani, A., Srivastava, A., Vieira, N.A., Slone, S. and Salmerón, M., 2022. Are soybean models ready for climate change food impact assessments? *European Journal of Agronomy*, 135.
- Kummu, M., Heino, M., Taka, M., Varis, O. and Viviroli, D., 2021. Climate change risks pushing one-third of global food production outside the safe climatic space. *One Earth*, 4(5): 720-729.
- Lal, R., 2006. Enhancing crop yields in the developing countries through restoration of the soil organic carbon pool in agricultural lands. *Land Degradation & Development*, 17(2): 197-209.
- Lam, H.-M., Remais, J., Fung, M.-C., Xu, L. and Sun, S.S.-M., 2013. Food supply and food safety issues in China. *The Lancet*, 381(9882): 2044-2053.
- Lange, S., 2019. Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0). *Geoscientific Model Development*, 12(7): 3055-3070.
- Lange, S., 2021. ISIMIP3BASD v2. 5.0. Zenodo, 10.
- Le, T., Ha, K.J. and Bae, D.H., 2021. Increasing Causal Effects of El Niño–Southern Oscillation on the Future Carbon Cycle of Terrestrial Ecosystems. *Geophysical Research Letters*, 48(24).
- Lee, J., Planton, Y.Y., Gleckler, P.J., Sperber, K.R., Guilyardi, E., Wittenberg, A.T., McPhaden, M.J. and Pallotta, G., 2021. Robust Evaluation of ENSO in Climate Models: How Many Ensemble Members Are Needed? *Geophysical Research Letters*, 48(20).
- Leng, G. and Hall, J., 2019. Crop yield sensitivity of global major agricultural countries to droughts and the projected changes in the future. *Sci Total Environ*, 654: 811-821.
- Leng, G. and Hall, J.W., 2020. Predicting spatial and temporal variability in crop yields: an inter-comparison of machine learning, regression and process-based models. *Environ Res Lett*, 15(4).
- Lesk, C., Anderson, W., Rigden, A., Coast, O., Jägermeyr, J., McDermid, S., Davis, K.F. and Konar, M., 2022. Compound heat and moisture extreme impacts on global crop yields under climate change. *Nature Reviews Earth & Environment*, 3(12): 872-889.

- Lesk, C., Coffel, E. and Horton, R., 2020. Net benefits to US soy and maize yields from intensifying hourly rainfall. *Nature Climate Change*, 10(9): 819-822.
- Lesk, C., Coffel, E., Winter, J., Ray, D., Zscheischler, J., Seneviratne, S.I. and Horton, R., 2021. Stronger temperature–moisture couplings exacerbate the impact of climate warming on global crop yields. *Nature Food*, 2(9): 683-691.
- Li, A., Liang, S., Wang, A. and Qin, J., 2007. Estimating crop yield from multi-temporal satellite data using multivariate regression and neural network techniques. *Photogrammetric Engineering & Remote Sensing*, 73(10): 1149-1157.
- Li, L., Wang, B., Feng, P., Li Liu, D., He, Q., Zhang, Y., Wang, Y., Li, S., Lu, X., Yue, C., Li, Y., He, J., Feng, H., Yang, G. and Yu, Q., 2022a. Developing machine learning models with multi-source environmental data to predict wheat yield in China. *Computers and Electronics in Agriculture*, 194.
- Li, L., Wang, B., Feng, P., Wang, H., He, Q., Wang, Y., Liu, D.L., Li, Y., He, J., Feng, H., Yang, G. and Yu, Q., 2021a. Crop yield forecasting and associated optimum lead time analysis based on multi-source environmental data across China. *Agricultural and Forest Meteorology*, 308-309.
- Li, L., Yao, N., Li, Y., Liu, D.L., Wang, B. and Ayantobo, O.O., 2019a. Future projections of extreme temperature events in different sub-regions of China. *Atmospheric Research*, 217: 150-164.
- Li, L., Yao, N., Liu, D.L., Song, S., Lin, H., Chen, X. and Li, Y., 2019b. Historical and future projected frequency of extreme precipitation indicators using the optimized cumulative distribution functions in China. *Journal of Hydrology*, 579.
- Li, L., Zou, Y., Li, Y., Lin, H., Liu, D.L., Wang, B., Yao, N. and Song, S., 2020a. Trends, change points and spatial variability in extreme precipitation events from 1961 to 2017 in China. *Hydrology Research*.
- Li, N., Lin, H., Wang, T., Li, Y., Liu, Y., Chen, X. and Hu, X., 2020b. Impact of climate change on cotton growth and yields in Xinjiang, China. *Field Crops Research*, 247.
- Li, P., Huang, Q., Huang, S., Leng, G., Peng, J., Wang, H., Zheng, X., Li, Y. and Fang, W., 2022b. Various maize yield losses and their dynamics triggered by drought thresholds based on Copula-Bayesian conditional probabilities. *Agricultural Water Management*, 261.
- Li, Y., Guan, K., Schnitkey, G.D., DeLucia, E. and Peng, B., 2019c. Excessive rainfall leads to maize yield loss of a comparable magnitude to extreme drought in the United States. *Glob Chang Biol*, 25(7): 2325-2337.
- Li, Y., Guan, K., Yu, A., Peng, B., Zhao, L., Li, B. and Peng, J., 2019d. Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the U.S. *Field Crops Research*, 234: 55-65.
- Li, Y., Zhou, Q., Zhou, J., Zhang, G., Chen, C. and Wang, J., 2014. Assimilating remote sensing information into a coupled hydrology-crop growth model to estimate regional maize yield in arid regions. *Ecological modelling*, 291: 15-27.
- Li, Z., Zhang, Z. and Zhang, L., 2021b. Improving regional wheat drought risk assessment for insurance application by integrating scenario-driven crop model, machine learning, and satellite data. *Agricultural Systems*, 191.
- Liakos, K.G., Busato, P., Moshou, D., Pearson, S. and Bochtis, D., 2018. Machine Learning in Agriculture: A Review. *Sensors (Basel)*, 18(8).

- Lin, H., Li, N., Li, Y., Liu, H., Liu, J., Li, L., Feng, P., Liu, D. and Liu, C., 2021. Quantitative Analysis of Winter Wheat Growth and Yields Responding to Climate Change in Xinjiang, China. *Water*, 13(24).
- Liu, A. and Zhao, S., 2018. High-performance target tracking scheme with low prediction precision requirement in WSNs. *International Journal of Ad Hoc and Ubiquitous Computing*, 29(4): 270-289.
- Liu, B., Asseng, S., Müller, C., Ewert, F., Elliott, J. and Lobell, D.B., 2016. Similar estimates of temperature impacts on global wheat yield by three independent methods. *Nature Climate Change*, 6(12): 1130-1136.
- Liu, B., Asseng, S., Wang, A., Wang, S., Tang, L., Cao, W., Zhu, Y. and Liu, L., 2017a. Modelling the effects of post-heading heat stress on biomass growth of winter wheat. *Agricultural and Forest Meteorology*, 247: 476-490.
- Liu, D., Zhou, C., Keesing, J.K., Serrano, O., Werner, A., Fang, Y., Chen, Y., Masque, P., Kinloch, J., Sadekov, A. and Du, Y., 2022. Wildfires enhance phytoplankton production in tropical oceans. *Nat Commun*, 13(1): 1348.
- Liu, D.L. and Zuo, H., 2012. Statistical downscaling of daily climate variables for climate change impact assessment over New South Wales, Australia. *Climatic Change*, 115(3-4): 629-666.
- Liu, H.-l., Liu, H.-b., Lei, Q.-l., Zhai, L.-m., Wang, H.-y., Zhang, J.-z., Zhu, Y.-p., Liu, S.-p., Li, S.-j., Zhang, J.-s. and Liu, X.-x., 2017b. Using the DSSAT model to simulate wheat yield and soil organic carbon under a wheat-maize cropping system in the North China Plain. *Journal of Integrative Agriculture*, 16(10): 2300-2307.
- Liu, K., Harrison, M.T., Yan, H., Liu, D.L., Meinke, H., Hoogenboom, G., Wang, B., Peng, B., Guan, K., Jaegermeyr, J., Wang, E., Zhang, F., Yin, X., Archontoulis, S., Nie, L., Badea, A., Man, J., Wallach, D., Zhao, J., Benjumea, A.B., Fahad, S., Tian, X., Wang, W., Tao, F., Zhang, Z., Rötter, R., Yuan, Y., Zhu, M., Dai, P., Nie, J., Yang, Y., Zhang, Y. and Zhou, M., 2023. Silver lining to a climate crisis in multiple prospects for alleviating crop waterlogging under future climates. *Nature Communications*, 14(1).
- Liu, L., Guan, L. and Liu, X., 2017c. Directly estimating diurnal changes in GPP for C3 and C4 crops using far-red sun-induced chlorophyll fluorescence. *Agricultural and Forest Meteorology*, 232: 1-9.
- Liu, W., Ye, T., Jägermeyr, J., Müller, C., Chen, S., Liu, X. and Shi, P., 2021a. Future climate change significantly alters interannual wheat yield variability over half of harvested areas. *Environmental Research Letters*, 16(9).
- Liu, X., Ferguson, R.B., Zheng, H., Cao, Q., Tian, Y., Cao, W. and Zhu, Y., 2017d. Using an Active-Optical Sensor to Develop an Optimal NDVI Dynamic Model for High-Yield Rice Production (Yangtze, China). *Sensors (Basel)*, 17(4).
- Liu, Y., Kumar, M., Katul, G.G. and Porporato, A., 2019. Reduced resilience as an early warning signal of forest mortality. *Nature Climate Change*, 9(11): 880-885.
- Liu, Z., Ying, H., Chen, M., Bai, J., Xue, Y., Yin, Y., Batchelor, W.D., Yang, Y., Bai, Z., Du, M., Guo, Y., Zhang, Q., Cui, Z., Zhang, F. and Dou, Z., 2021b. Optimization of China's maize and soy production can ensure feed sufficiency at lower nitrogen and carbon footprints. *Nature Food*, 2(6): 426-433.
- Lobell, D.B., Cahill, K.N. and Field, C.B., 2007. Historical effects of temperature and precipitation on California crop yields. *Climatic Change*, 81(2): 187-203.
- Lu, C., Yu, Z., Zhang, J., Cao, P., Tian, H. and Nevison, C., 2022. Century-long changes and drivers of soil nitrous oxide (N<sub>2</sub>O) emissions across the contiguous United States. *Glob Chang Biol*, 28(7):

2505-2524.

- Lyon, C., Saupe, E.E., Smith, C.J., Hill, D.J., Beckerman, A.P., Stringer, L.C., Marchant, R., McKay, J., Burke, A., O'Higgins, P., Dunhill, A.M., Allen, B.J., Riel-Salvatore, J. and Aze, T., 2022. Climate change research and action must look beyond 2100. *Glob Chang Biol*, 28(2): 349-361.
- Maiorano, A., Martre, P., Asseng, S., Ewert, F., Müller, C., Rötter, R.P., Ruane, A.C., Semenov, M.A., Wallach, D., Wang, E., Alderman, P.D., Kassie, B.T., Biernath, C., Basso, B., Cammarano, D., Challinor, A.J., Doltra, J., Dumont, B., Rezaei, E.E., Gayler, S., Kersebaum, K.C., Kimball, B.A., Koehler, A.-K., Liu, B., O'Leary, G.J., Olesen, J.E., Ottman, M.J., Priesack, E., Reynolds, M., Stratonovitch, P., Streck, T., Thorburn, P.J., Waha, K., Wall, G.W., White, J.W., Zhao, Z. and Zhu, Y., 2017. Crop model improvement reduces the uncertainty of the response to temperature of multi-model ensembles. *Field Crops Research*, 202: 5-20.
- Majumder, B., Mandal, B., Bandyopadhyay, P., Gangopadhyay, A., Mani, P., Kundu, A. and Mazumdar, D., 2008. Organic amendments influence soil organic carbon pools and rice–wheat productivity. *Soil science society of America journal*, 72(3): 775-785.
- Maya Gopal, P.S. and Bhargavi, R., 2019. Performance Evaluation of Best Feature Subsets for Crop Yield Prediction Using Machine Learning Algorithms. *Applied Artificial Intelligence*, 33(7): 621-642.
- Mbow, C., Rosenzweig, C., Barioni, L.G., Benton, T.G., Herrero, M., Krishnapillai, M., Liwenga, E., Pradhan, P., Rivera-Ferre, M. and Sapkota, T., 2019. Food security, Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security and greenhouse gas fluxes in terrestrial ecosystems. IPCC.
- McCullough, E.B., Quinn, J.D. and Simons, A.M., 2022. Profitability of climate-smart soil fertility investment varies widely across sub-Saharan Africa. *Nature Food*, 3(4): 275-285.
- McDonald, A.J., Balwinder, S., Keil, A., Srivastava, A., Craufurd, P., Kishore, A., Kumar, V., Paudel, G., Singh, S., Singh, A.K., Sohane, R.K. and Malik, R.K., 2022. Time management governs climate resilience and productivity in the coupled rice–wheat cropping systems of eastern India. *Nature Food*, 3(7): 542-551.
- McKenna, C.M. and Maycock, A.C., 2021. Sources of Uncertainty in Multimodel Large Ensemble Projections of the Winter North Atlantic Oscillation. *Geophysical Research Letters*, 48(14).
- McKenna, C.M. and Maycock, A.C., 2022. The Role of the North Atlantic Oscillation for Projections of Winter Mean Precipitation in Europe. *Geophysical Research Letters*, 49(19).
- McPhaden, M.J., Santoso, A. and Cai, W., 2020. *El Niño Southern Oscillation in a changing climate*, 253. John Wiley & Sons.
- Meng, W., Tao, F. and Shi, W., 2014. Corn yield forecasting in northeast china using remotely sensed spectral indices and crop phenology metrics. *Journal of Integrative Agriculture*, 13(7): 1538-1545.
- Minoli, S., Jagermeyr, J., Asseng, S., Urfels, A. and Muller, C., 2022. Global crop yields can be lifted by timely adaptation of growing periods to climate change. *Nat Commun*, 13(1): 7079.
- Moldenhauer, K. and Slaton, N., 2001. Rice growth and development. *Rice production handbook*, 192: 7-14.
- Monfreda, C., Ramankutty, N. and Foley, J.A., 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles*, 22(1): n/a-n/a.

- Morim, J., Hemer, M., Wang, X.L., Cartwright, N., Trenham, C., Semedo, A., Young, I., Bricheno, L., Camus, P., Casas-Prat, M., Erikson, L., Mentaschi, L., Mori, N., Shimura, T., Timmermans, B., Aarnes, O., Breivik, Ø., Behrens, A., Dobrynin, M., Menendez, M., Staneva, J., Wehner, M., Wolf, J., Kamranzad, B., Webb, A., Stopa, J. and Andutta, F., 2019. Robustness and uncertainties in global multivariate wind-wave climate projections. *Nature Climate Change*, 9(9): 711-718.
- Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M. and Hoek, S., 2017. Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, 10(4): 1403-1422.
- Müller, C., Elliott, J., Kelly, D., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Hoek, S. and Izaurrealde, R.C., 2019. The Global Gridded Crop Model Intercomparison phase 1 simulation dataset. *Scientific data*, 6(1): 1-22.
- Müller, C., Franke, J., Jägermeyr, J., Ruane, A.C., Elliott, J., Moyer, E., Heinke, J., Falloon, P.D., Folberth, C., Francois, L., Hank, T., Izaurrealde, R.C., Jacquemin, I., Liu, W., Olin, S., Pugh, T.A.M., Williams, K. and Zabel, F., 2021. Exploring uncertainties in global crop yield projections in a large ensemble of crop models and CMIP5 and CMIP6 climate scenarios. *Environmental Research Letters*, 16(3).
- Naimi, S., Ayoubi, S., Demattê, J.A., Zeraatpisheh, M., Amorim, M.T.A. and Mello, F.A.d.O., 2021. Spatial prediction of soil surface properties in an arid region using synthetic soil image and machine learning. *Geocarto International*: 1-24.
- Nilsson, J. and Lejenäs, H., 2011. The 2010 Crafoord Prize awarded to Walter Munk. *Tellus A: Dynamic Meteorology and Oceanography*, 63(2).
- Norouzi, M., Ayoubi, S., Jalalian, A., Khademi, H. and Dehghani, A., 2010. Predicting rainfed wheat quality and quantity by artificial neural network using terrain and soil characteristics. *Acta Agriculturae Scandinavica Section B—Soil Plant Science*, 60(4): 341-352.
- Nouri, N., Devineni, N., Were, V. and Khanbilvardi, R., 2021. Explaining the trends and variability in the United States tornado records using climate teleconnections and shifts in observational practices. *Sci Rep*, 11(1): 1741.
- Obermeier, W.A., Lehnert, L.W., Kammann, C.I., Müller, C., Grünhage, L., Luterbacher, J., Erbs, M., Moser, G., Seibert, R., Yuan, N. and Bendix, J., 2016. Reduced CO<sub>2</sub> fertilization effect in temperate C3 grasslands under more extreme weather conditions. *Nature Climate Change*, 7(2): 137-141.
- Otteson, B.N., Mergoum, M., Ransom, J. and Schatz, B., 2008. Tiller contribution to spring wheat yield under varying seeding and nitrogen management. *Agronomy journal*, 100(2): 406-413.
- Ottman, M., Kimball, B., White, J. and Wall, G., 2012. Wheat growth response to increased temperature from varied planting dates and supplemental infrared heating. *Agronomy Journal*, 104(1): 7-16.
- Peng, B. and Guan, K., 2021. Harmonizing climate-smart and sustainable agriculture. *Nature Food*, 2(11): 853-854.
- Peng, B., Guan, K., Tang, J., Ainsworth, E.A., Asseng, S., Bernacchi, C.J., Cooper, M., Delucia, E.H., Elliott, J.W., Ewert, F., Grant, R.F., Gustafson, D.I., Hammer, G.L., Jin, Z., Jones, J.W., Kimm, H., Lawrence, D.M., Li, Y., Lombardozzi, D.L., Marshall-Colon, A., Messina, C.D., Ort, D.R., Schnable, J.C., Vallejos, C.E., Wu, A., Yin, X. and Zhou, W., 2020b. Towards a multiscale crop modelling framework for climate change adaptation assessment. *Nat Plants*, 6(4): 338-348.



- Peng, B., Guan, K., Zhou, W., Jiang, C., Frankenberg, C., Sun, Y., He, L. and Köhler, P., 2020a. Assessing the benefit of satellite-based Solar-Induced Chlorophyll Fluorescence in crop yield prediction. *International Journal of Applied Earth Observation and Geoinformation*, 90.
- Perry, E.D., Yu, J. and Tack, J., 2020. Using insurance data to quantify the multidimensional impacts of warming temperatures on yield risk. *Nat Commun*, 11(1): 4542.
- Phillips, N. and Nogrady, B., 2020. The race to decipher how climate change influenced Australia's record fires. *Nature*, 577(7791): 610-613.
- Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., Zhou, L., Liu, H., Ma, Y., Ding, Y., Friedlingstein, P., Liu, C., Tan, K., Yu, Y., Zhang, T. and Fang, J., 2010. The impacts of climate change on water resources and agriculture in China. *Nature*, 467(7311): 43-51.
- Prado, R. and West, M., 2010. *Time series: modeling, computation, and inference*. Chapman and Hall/CRC.
- Puma, M.J., Bose, S., Chon, S.Y. and Cook, B.I., 2015. Assessing the evolving fragility of the global food system. *Environmental Research Letters*, 10(2).
- Rattalino Edreira, J.I., Andrade, J.F., Cassman, K.G., van Ittersum, M.K., van Loon, M.P. and Grassini, P., 2021. Spatial frameworks for robust estimation of yield gaps. *Nature Food*, 2(10): 773-779.
- Ray, D.K., Gerber, J.S., MacDonald, G.K. and West, P.C., 2015. Climate variation explains a third of global crop yield variability. *Nat Commun*, 6: 5989.
- Rayner, N.A., 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108(D14).
- Rehfeldt, G.E., Crookston, N.L., Sáenz-Romero, C. and Campbell, E.M., 2012. North American vegetation model for land - use planning in a changing climate: A solution to large classification problems. *Ecological Applications*, 22(1): 119-141.
- Roden, J.S. and Pearcy, R.W., 1993. Effect of leaf flutter on the light environment of poplars. *Oecologia*, 93(2): 201-207.
- Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Muller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T.A., Schmid, E., Stehfest, E., Yang, H. and Jones, J.W., 2014a. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proc Natl Acad Sci U S A*, 111(9): 3268-73.
- Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T.A.M., Schmid, E., Stehfest, E., Yang, H. and Jones, J.W., 2014b. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences*, 111(9): 3268.
- Rossato, L., Alvalá, R.C.d.S., Marengo, J.A., Zeri, M., Cunha, A.P.M.d.A., Pires, L.B.M. and Barbosa, H.A., 2017. Impact of Soil Moisture on Crop Yields over Brazilian Semiarid. *Frontiers in Environmental Science*, 5.
- Royo, C., Villegas, D., Rharrabti, Y., Blanco, R., Martos, V. and García del Moral, L., 2006. Grain growth and yield formation of durum wheat grown at contrasting latitudes and water regimes in a Mediterranean environment. *Cereal Research Communications*, 34(2-3): 1021-1028.
- Ruane, A.C. and McDermid, S.P., 2017. Selection of a representative subset of global climate models that captures the profile of regional changes for integrated climate impacts assessment. *Earth Perspectives*,

4(1).

- Ryu, Y., Berry, J.A. and Baldocchi, D.D., 2019. What is global photosynthesis? History, uncertainties and opportunities. *Remote sensing of environment*, 223: 95-114.
- Saji, N. and Yamagata, T., 2003. Possible impacts of Indian Ocean dipole mode events on global climate. *Climate Research*, 25(2): 151-169.
- Saji, N., Goswami, B.N., Vinayachandran, P. and Yamagata, T., 1999. A dipole mode in the tropical Indian Ocean. *Nature*, 401(6751): 360-363.
- Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Muller, C., Pugh, T.A., Rolinski, S., Schaphoff, S., Schmid, E., Wang, X., Schlenker, W. and Frieler, K., 2017a. Consistent negative response of US crops to high temperatures in observations and crop models. *Nat Commun*, 8: 13931.
- Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N. and Müller, C., 2017b. Consistent negative response of US crops to high temperatures in observations and crop models. *Nature communications*, 8(1): 1-9.
- Schillerberg, T.A. and Tian, D., 2020. Changes of crop failure risks in the United States associated with large-scale climate oscillations in the Atlantic and Pacific Oceans. *Environmental Research Letters*, 15(6).
- Schillerberg, T.A., Tian, D. and Miao, R., 2019. Spatiotemporal patterns of maize and winter wheat yields in the United States: Predictability and impact from climate oscillations. *Agricultural and Forest Meteorology*, 275: 208-222.
- Schwalbert, R.A., Amado, T., Corassa, G., Pott, L.P., Prasad, P.V.V. and Ciampitti, I.A., 2020. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology*, 284.
- Shew, A.M., Tack, J.B., Nalley, L.L. and Chaminuka, P., 2020. Yield reduction under climate warming varies among wheat cultivars in South Africa. *Nat Commun*, 11(1): 4408.
- Shuai, J., Zhang, Z., Tao, F. and Shi, P., 2016. How ENSO affects maize yields in China: understanding the impact mechanisms using a process-based crop model. *International Journal of Climatology*, 36(1): 424-438.
- Siebert, S., Webber, H., Zhao, G. and Ewert, F., 2017. Heat stress is overestimated in climate impact studies for irrigated agriculture. *Environmental Research Letters*, 12(5).
- Silvestro, P., Pignatti, S., Pascucci, S., Yang, H., Li, Z., Yang, G., Huang, W. and Casa, R., 2017. Estimating Wheat Yield in China at the Field and District Scale from the Assimilation of Satellite Data into the Aquacrop and Simple Algorithm for Yield (SAFY) Models. *Remote Sensing*, 9(5).
- Singh, B., Sihag, P. and Singh, K., 2017. Modelling of impact of water quality on infiltration rate of soil by random forest regression. *Modeling Earth Systems Environment*, 3(3): 999-1004.
- Singh, J., Ashfaq, M., Skinner, C.B., Anderson, W.B., Mishra, V. and Singh, D., 2022. Enhanced risk of concurrent regional droughts with increased ENSO variability and warming. *Nature Climate Change*, 12(2): 163-170.
- Sippel, S., Reichstein, M., Ma, X., Mahecha, M.D., Lange, H., Flach, M. and Frank, D., 2018. Drought, Heat, and the Carbon Cycle: a Review. *Current Climate Change Reports*, 4(3): 266-286.

- Son, N., Chen, C., Chen, C., Minh, V. and Trung, N., 2014. A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation. *Agricultural and forest meteorology*, 197: 52-64.
- Song, L., Guanter, L., Guan, K., You, L., Huete, A., Ju, W. and Zhang, Y., 2018. Satellite sun-induced chlorophyll fluorescence detects early response of winter wheat to heat stress in the Indian Indo-Gangetic Plains. *Glob Chang Biol*, 24(9): 4023-4037.
- Stephenson, D.B., Pavan, V., Collins, M., Junge, M.M. and Quadrelli, R., 2006. North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment. *Climate Dynamics*, 27(4): 401-420.
- Stephenson, N., 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. *Journal of biogeography*, 25(5): 855-870.
- Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P.M., 2013. The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Climate Change*.
- Sui, B., Feng, X., Tian, G., Hu, X., Shen, Q. and Guo, S., 2013. Optimizing nitrogen supply increases rice yield and nitrogen use efficiency by regulating yield formation factors. *Field Crops Research*, 150: 99-107.
- Sun, D.-W. and Woods, J., 1994. Low temperature moisture transfer characteristics of wheat in thin layers. *Transactions of the ASAE*, 37(6): 1919-1926.
- Swann, A.L., Hoffman, F.M., Koven, C.D. and Randerson, J.T., 2016. Plant responses to increasing CO<sub>2</sub> reduce estimates of climate impacts on drought severity. *Proc Natl Acad Sci U S A*, 113(36): 10019-24.
- Tang, T., Luo, J.J., Peng, K., Qi, L. and Tang, S., 2021. Over-projected Pacific warming and extreme El Nino frequency due to CMIP5 common biases. *Natl Sci Rev*, 8(10): nwab056.
- Tao, F., Palosuo, T., Rötter, R.P., Díaz-Ambrona, C.G.H., Inés Mínguez, M., Semenov, M.A., Kersebaum, K.C., Cammarano, D., Specka, X., Nendel, C., Srivastava, A.K., Ewert, F., Padovan, G., Ferrise, R., Martre, P., Rodríguez, L., Ruiz-Ramos, M., Gaiser, T., Höhn, J.G., Salo, T., Dibari, C. and Schulman, A.H., 2020. Why do crop models diverge substantially in climate impact projections? A comprehensive analysis based on eight barley crop models. *Agricultural and Forest Meteorology*, 281.
- Tao, F., Rötter, R.P., Palosuo, T., Díaz-Ambrona, C.G.H., Mínguez, M.I., Semenov, M.A., Kersebaum, K.C., Nendel, C., Cammarano, D., Hoffmann, H., Ewert, F., Dambreville, A., Martre, P., Rodríguez, L., Ruiz-Ramos, M., Gaiser, T., Höhn, J.G., Salo, T., Ferrise, R., Bindi, M. and Schulman, A.H., 2017. Designing future barley ideotypes using a crop model ensemble. *European Journal of Agronomy*, 82: 144-162.
- Tao, F., Rotter, R.P., Palosuo, T., Gregorio Hernandez Diaz-Ambrona, C., Mínguez, M.I., Semenov, M.A., Kersebaum, K.C., Nendel, C., Specka, X., Hoffmann, H., Ewert, F., Dambreville, A., Martre, P., Rodríguez, L., Ruiz-Ramos, M., Gaiser, T., Hohn, J.G., Salo, T., Ferrise, R., Bindi, M., Cammarano, D. and Schulman, A.H., 2018. Contribution of crop model structure, parameters and climate projections to uncertainty in climate change impact assessments. *Glob Chang Biol*, 24(3): 1291-1307.
- Tao, F., Yokozawa, M., Liu, J. and Zhang, Z., 2008. Climate–crop yield relationships at provincial scales in China and the impacts of recent climate trends. *Climate Research*, 38: 83-94.

- Tian, H., Yang, J., Xu, R., Lu, C., Canadell, J.G., Davidson, E.A., Jackson, R.B., Arneeth, A., Chang, J., Ciais, P., Gerber, S., Ito, A., Joos, F., Lienert, S., Messina, P., Olin, S., Pan, S., Peng, C., Saikawa, E., Thompson, R.L., Vuichard, N., Winiwarter, W., Zaehle, S. and Zhang, B., 2019. Global soil nitrous oxide emissions since the preindustrial era estimated by an ensemble of terrestrial biosphere models: Magnitude, attribution, and uncertainty. *Glob Chang Biol*, 25(2): 640-659.
- Tilman, D., Balzer, C., Hill, J. and Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. *Proc Natl Acad Sci U S A*, 108(50): 20260-4.
- Vecchi, G.A. and Soden, B.J., 2007. Effect of remote sea surface temperature change on tropical cyclone potential intensity. *Nature*, 450(7172): 1066-70.
- Vetter, T., Reinhardt, J., Flörke, M., van Griensven, A., Hattermann, F., Huang, S., Koch, H., Pechlivanidis, I.G., Plötner, S., Seidou, O., Su, B., Vervoort, R.W. and Krysanova, V., 2017. Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins. *Climatic Change*, 141(3): 419-433.
- Vicente-Serrano, S.M., Beguería, S. and López-Moreno, J.I., 2010. A multiscale drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, 23(7): 1696-1718.
- Wallach, D., Palosuo, T., Thorburn, P., Gourdain, E., Asseng, S., Basso, B., Buis, S., Crout, N., Dibari, C. and Dumont, B., 2021. How well do crop modeling groups predict wheat phenology, given calibration data from the target population? *European Journal of Agronomy*, 124: 126195.
- Wang, B., Feng, P., Liu, D.L., O'Leary, G.J., Macadam, I., Waters, C., Asseng, S., Cowie, A., Jiang, T., Xiao, D., Ruan, H., He, J. and Yu, Q., 2020b. Sources of uncertainty for wheat yield projections under future climate are site-specific. *Nature Food*, 1(11): 720-728.
- Wang, B., Feng, P., Waters, C., Cleverly, J., Liu, D.L. and Yu, Q., 2020a. Quantifying the impacts of pre-occurred ENSO signals on wheat yield variation using machine learning in Australia. *Agricultural and Forest Meteorology*, 291.
- Wang, B., Liu, D.L., Macadam, I., Alexander, L.V., Abramowitz, G. and Yu, Q., 2016a. Multi-model ensemble projections of future extreme temperature change using a statistical downscaling method in south eastern Australia. *Climatic Change*, 138(1-2): 85-98.
- Wang, B., Liu, D.L., O'Leary, G.J., Asseng, S., Macadam, I., Lines-Kelly, R., Yang, X., Clark, A., Crean, J., Sides, T., Xing, H., Mi, C. and Yu, Q., 2018a. Australian wheat production expected to decrease by the late 21st century. *Global Change Biology*, 24(6): 2403-2415.
- Wang, B., Waters, C., Orgill, S., Cowie, A., Clark, A., Li Liu, D., Simpson, M., McGowen, I. and Sides, T., 2018b. Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. *Ecological Indicators*, 88: 425-438.
- Wang, B., Wu, R. and Fu, X., 2000. Pacific–East Asian teleconnection: how does ENSO affect East Asian climate? *Journal of Climate*, 13(9): 1517-1536.
- Wang, C., Wang, X., Jin, Z., Müller, C., Pugh, T.A.M., Chen, A., Wang, T., Huang, L., Zhang, Y., Li, L.X.Z. and Piao, S., 2021a. Occurrence of crop pests and diseases has largely increased in China since 1970. *Nature Food*, 3(1): 57-65.
- Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., Rotter, R.P., Kimball, B.A., Ottman, M.J., Wall, G.W.,

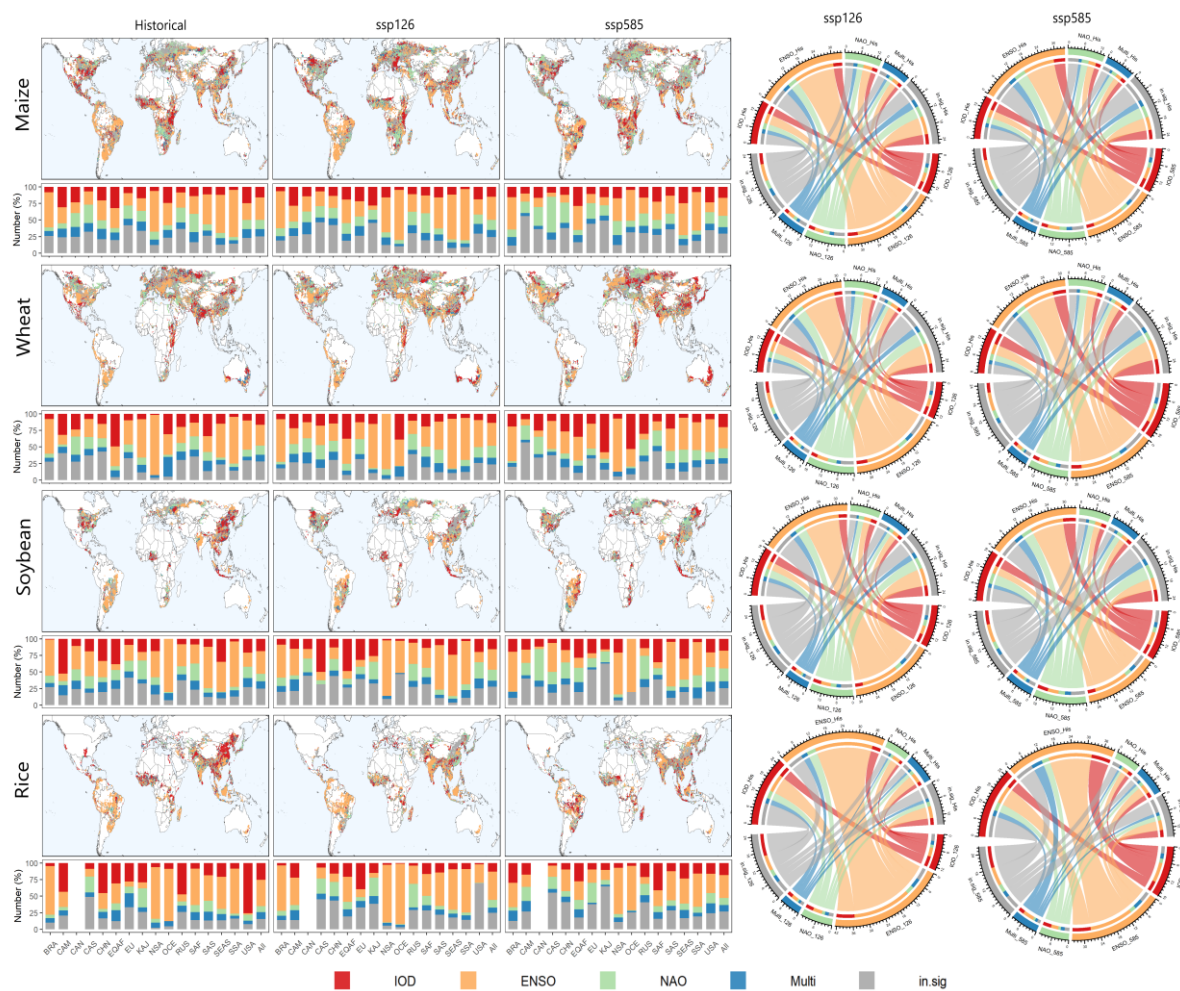
- White, J.W., Reynolds, M.P., Alderman, P.D., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A.J., De Sanctis, G., Doltra, J., Dumont, B., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.K., Liu, L., Muller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ripoche, D., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stockle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wallach, D., Wang, Z., Wolf, J., Zhu, Y. and Asseng, S., 2017b. The uncertainty of crop yield projections is reduced by improved temperature response functions. *Nat Plants*, 3: 17102.
- Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., Rötter, R.P., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A.J., De Sanctis, G., Doltra, J., Dumont, B., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.-K., Liu, L., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ripoche, D., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wallach, D., Wang, Z., Wolf, J., Zhu, Y. and Asseng, S., 2017a. The uncertainty of crop yield projections is reduced by improved temperature response functions. *Nature Plants*, 3(8): 17102.
- Wang, H., Vicente-Serrano, S.M., Tao, F., Zhang, X., Wang, P., Zhang, C., Chen, Y., Zhu, D. and Kenawy, E., 2016b. Monitoring winter wheat drought threat in Northern China using multiple climate-based drought indices and soil moisture during 2000–2013. *Agricultural and Forest Meteorology*, 228: 1-12.
- Wang, H.M., Chen, J., Xu, C.Y., Zhang, J. and Chen, H., 2020c. A Framework to Quantify the Uncertainty Contribution of GCMs Over Multiple Sources in Hydrological Impacts of Climate Change. *Earth's Future*, 8(8).
- Wang, J., Wang, M., Kim, J.S., Joiner, J., Zeng, N., Jiang, F., Wang, H., He, W., Wu, M., Chen, T., Ju, W. and Chen, J.M., 2021b. Modulation of Land Photosynthesis by the Indian Ocean Dipole: Satellite - Based Observations and CMIP6 Future Projections. *Earth's Future*, 9(4).
- Wang, J., Zhang, J., Bai, Y., Zhang, S., Yang, S. and Yao, F., 2020d. Integrating remote sensing-based process model with environmental zonation scheme to estimate rice yield gap in Northeast China. *Field Crops Research*, 246: 107682.
- Wang, L., Li, Y., Li, M., Li, L., Liu, F., Liu, D.L. and Pulatov, B., 2022a. Projection of precipitation extremes in China's mainland based on the statistical downscaled data from 27 GCMs in CMIP6. *Atmospheric Research*, 280.
- Wang, X., Huang, J., Feng, Q. and Yin, D., 2020g. Winter Wheat Yield Prediction at County Level and Uncertainty Analysis in Main Wheat-Producing Regions of China with Deep Learning Approaches. *Remote Sensing*, 12(11).
- Wang, X., Muller, C., Elliot, J., Mueller, N.D., Ciais, P., Jagermeyr, J., Gerber, J., Dumas, P., Wang, C., Yang, H., Li, L., Deryng, D., Folberth, C., Liu, W., Makowski, D., Olin, S., Pugh, T.A.M., Reddy, A., Schmid, E., Jeong, S., Zhou, F. and Piao, S., 2021c. Global irrigation contribution to wheat and maize yield. *Nat Commun*, 12(1): 1235.
- Wang, X., Wang, S., Li, X., Chen, B., Wang, J., Huang, M. and Rahman, A., 2020f. Modelling rice yield with temperature optima of rice productivity derived from satellite NIRv in tropical monsoon area.

- Agricultural and Forest Meteorology, 294.
- Wang, X., Zhao, C., Müller, C., Wang, C., Ciaï, P., Janssens, I., Peñuelas, J., Asseng, S., Li, T., Elliott, J., Huang, Y., Li, L. and Piao, S., 2020e. Emergent constraint on crop yield response to warmer temperature from field experiments. *Nature Sustainability*, 3(11): 908-916.
- Wang, Y., Liu, Z., Yu, Q., Liu, L., Liu, X., Li, L., Jia, Q., Guo, C. and Lu, X., 2022b. Simulations of solar-induced chlorophyll fluorescence over crop canopies using the integrated APSIM model. *Computers and Electronics in Agriculture*, 203.
- Wheeler, T. and Braun, V., 2013. Climate change impacts on global food security. *Science*, 341: 508–513.
- Wheeler, T. and von Braun, J., 2013. Climate change impacts on global food security. *Science*, 341(6145): 508-13.
- Wu, A., Hammer, G.L., Doherty, A., von Caemmerer, S. and Farquhar, G.D., 2019. Quantifying impacts of enhancing photosynthesis on crop yield. *Nat Plants*, 5(4): 380-388.
- Xiao, L., Liu, L., Asseng, S., Xia, Y., Tang, L., Liu, B., Cao, W. and Zhu, Y., 2018. Estimating spring frost and its impact on yield across winter wheat in China. *Agricultural and Forest Meteorology*, 260-261: 154-164.
- Xiong, W., Asseng, S., Hoogenboom, G., Hernandez-Ochoa, I., Robertson, R., Sonder, K., Pequeno, D., Reynolds, M. and Gerard, B., 2019. Different uncertainty distribution between high and low latitudes in modelling warming impacts on wheat. *Nature Food*, 1(1): 63-69.
- Yang, X., Chen, F., Lin, X., Liu, Z., Zhang, H., Zhao, J., Li, K., Ye, Q., Li, Y., Lv, S., Yang, P., Wu, W., Li, Z., Lal, R. and Tang, H., 2015. Potential benefits of climate change for crop productivity in China. *Agricultural and Forest Meteorology*, 208: 76-84.
- Yao, N., Li, L., Feng, P., Feng, H., Li, L., Liu, Y., Jiang, K., Hu, X. and Li, Y., 2020. Projections of drought characteristics in China based on a standardized precipitation and evapotranspiration index and multiple GCMs. *Sci Total Environ*, 704: 135245.
- Yao, N., Li, Y., Dong, Q.g., Li, L., Peng, L. and Feng, H., 2019. Influence of the accuracy of reference crop evapotranspiration on drought monitoring using standardized precipitation evapotranspiration index in mainland China. *Land Degradation & Development*, 31(2): 266-282.
- Yao, N., Li, Y., Lei, T. and Peng, L., 2018. Drought evolution, severity and trends in mainland China over 1961–2013. *Science of The Total Environment*, 616-617: 73-89.
- Zabel, F., Müller, C., Elliott, J., Minoli, S., Jägermeyr, J., Schneider, J.M., Franke, J.A., Moyer, E., Dury, M., Francois, L., Folberth, C., Liu, W., Pugh, T.A.M., Olin, S., Rabin, S.S., Mauser, W., Hank, T., Ruane, A.C. and Asseng, S., 2021. Large potential for crop production adaptation depends on available future varieties. *Global Change Biology*.
- Zegada-Lizarazu, W., Wullschleger, S.D., Surendran Nair, S. and Monti, A., 2012. Crop Physiology, Switchgrass. *Green Energy and Technology*, pp. 55-86.
- Zeileis, A., Leisch, F., Kleiber, C. and Hornik, K., 2005. Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics*, 20(1): 99-121.
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S. and Finke, P., 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma*, 338: 445-452.



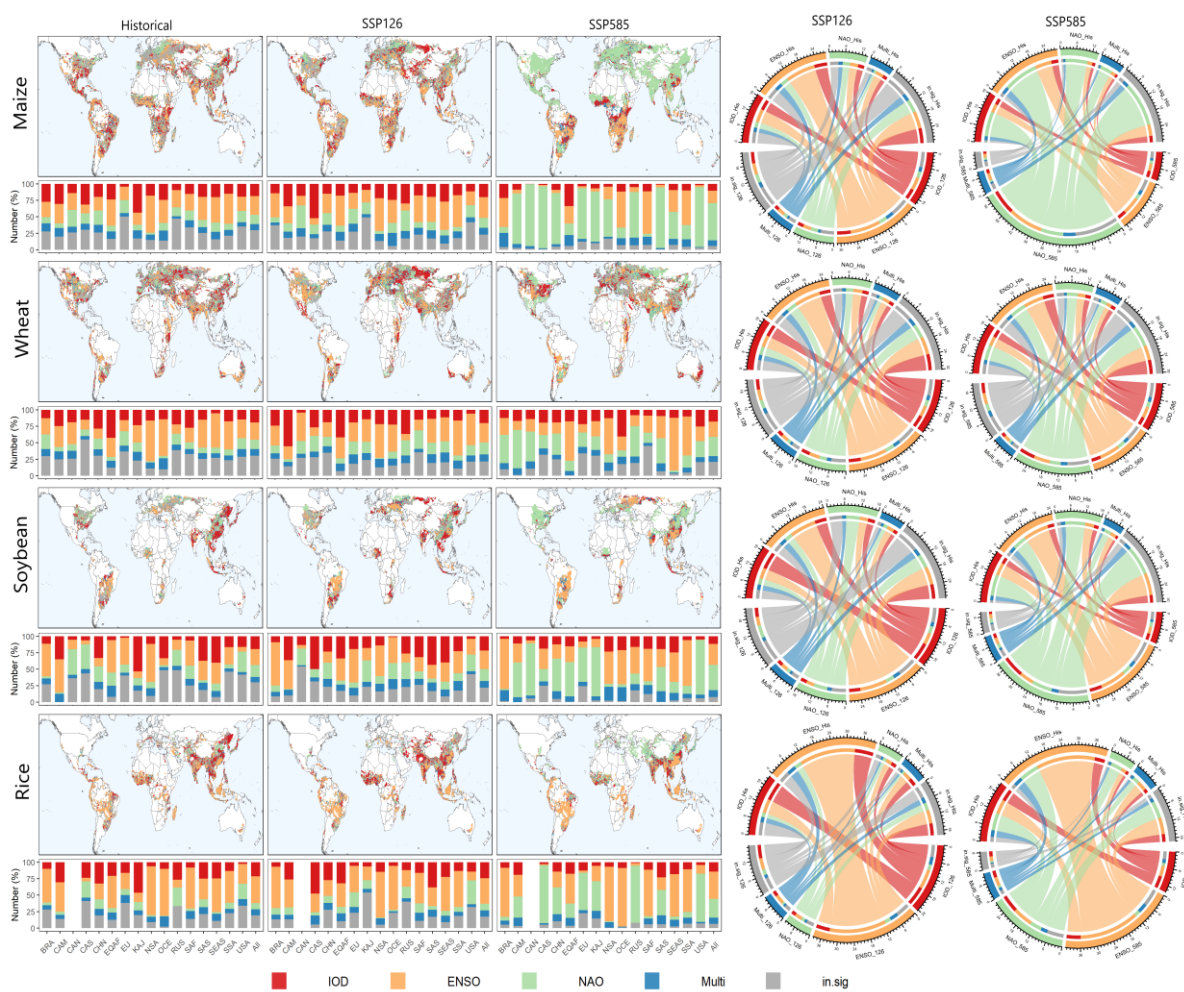
- Zeraatpisheh, M., Ayoubi, S., Mirbagheri, Z., Mosaddeghi, M.R. and Xu, M., 2021. Spatial prediction of soil aggregate stability and soil organic carbon in aggregate fractions using machine learning algorithms and environmental variables. *Geoderma Regional*, 27: e00440.
- Zhang, Y., Gentine, P., Luo, X., Lian, X., Liu, Y., Zhou, S., Michalak, A.M., Sun, W., Fisher, J.B., Piao, S. and Keenan, T.F., 2022. Increasing sensitivity of dryland vegetation greenness to precipitation due to rising atmospheric CO<sub>2</sub>. *Nat Commun*, 13(1): 4875.
- Zhang, Y., Joiner, J., Alemohammad, S.H., Zhou, S. and Gentine, P., 2018. A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks. *Biogeosciences*, 15(19): 5779-5800.
- Zhang, Y., Keenan, T.F. and Zhou, S., 2021. Exacerbated drought impacts on global ecosystems due to structural overshoot. *Nat Ecol Evol*, 5(11): 1490-1498.
- Zhao, C., Piao, S., Huang, Y., Wang, X., Ciais, P., Huang, M., Zeng, Z. and Peng, S., 2016a. Field warming experiments shed light on the wheat yield response to temperature in China. *Nat Commun*, 7: 13530.
- Zhao, C., Piao, S., Wang, X., Huang, Y., Ciais, P., Elliott, J., Huang, M., Janssens, I.A., Li, T., Lian, X., Liu, Y., Muller, C., Peng, S., Wang, T., Zeng, Z. and Penuelas, J., 2016b. Plausible rice yield losses under future climate warming. *Nat Plants*, 3: 16202.
- Zhao, X., Yu, B., Liu, Y., Chen, Z., Li, Q., Wang, C. and Wu, J., 2019. Estimation of poverty using random forest regression with multi-source data: A case study in Bangladesh. *Remote Sensing*, 11(4): 375.
- Zhao, Z., Wang, E., Kirkegaard, J.A. and Rebetzke, G.J., 2022. Novel wheat varieties facilitate deep sowing to beat the heat of changing climates. *Nature Climate Change*, 12(3): 291-296.
- Zhou, S., Williams, A.P., Berg, A.M., Cook, B.I., Zhang, Y., Hagemann, S., Lorenz, R., Seneviratne, S.I. and Gentine, P., 2019a. Land-atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity. *Proc Natl Acad Sci U S A*, 116(38): 18848-18853.
- Zhou, S., Zhang, Y., Park Williams, A. and Gentine, P., 2019b. Projected increases in intensity, frequency, and terrestrial carbon costs of compound drought and aridity events. *Sci Adv*, 5(1): eaau5740.

## 附录



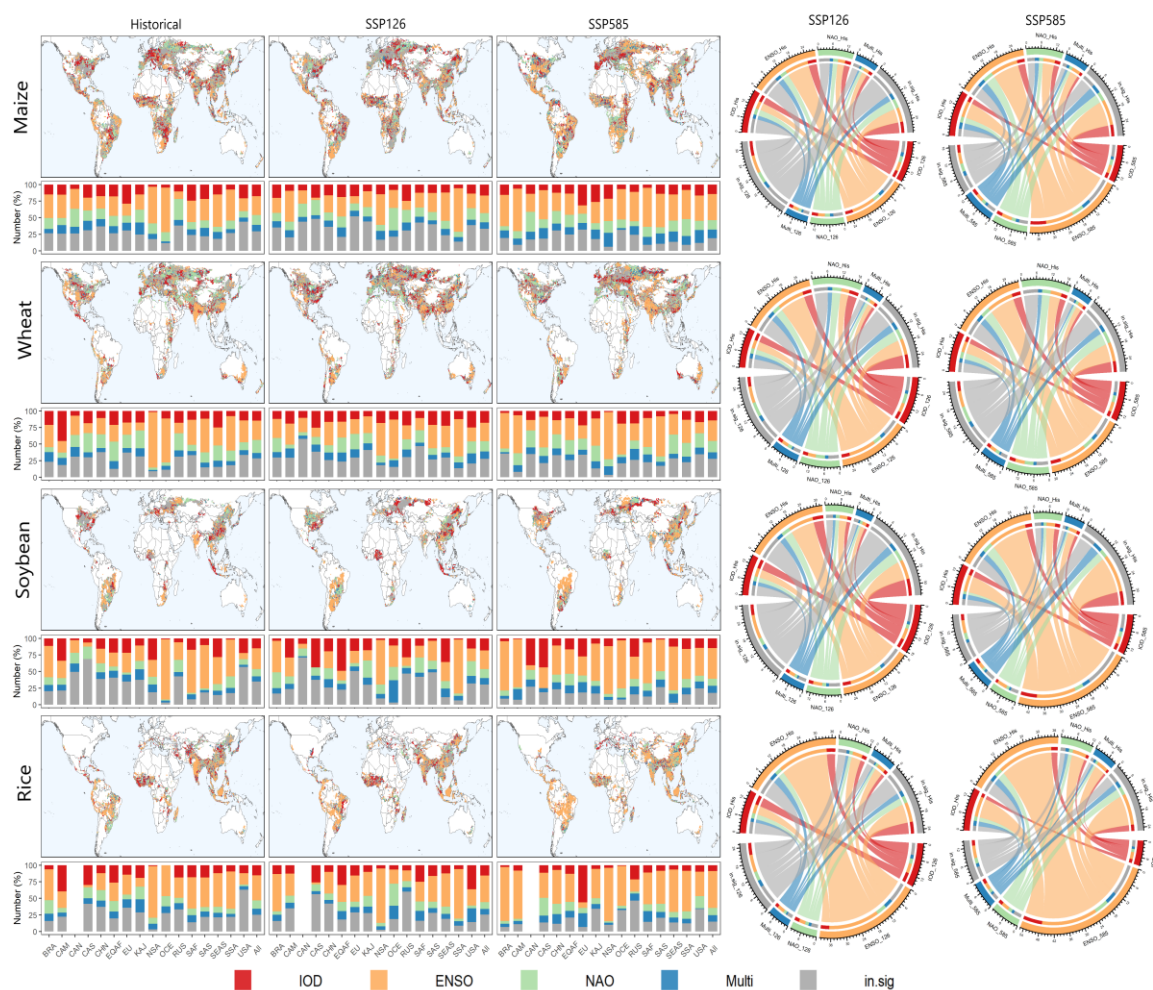
附图 1 基于 MRI-ESM2-0 气候模式，由 RF 模型确定的每个网格的玉米、小麦、大豆和水稻的主导大气环流指数

Fig. S1 Dominant climate oscillation indices of global maize, wheat, soybean and rice yield at each grid as identified by the RF model based on MRI-ESM2-0



附图 2 基于 IPSL-CM6A-LR 气候模式，由 RF 模型确定的每个网格的全球玉米和小麦产量的主导大气环流指数

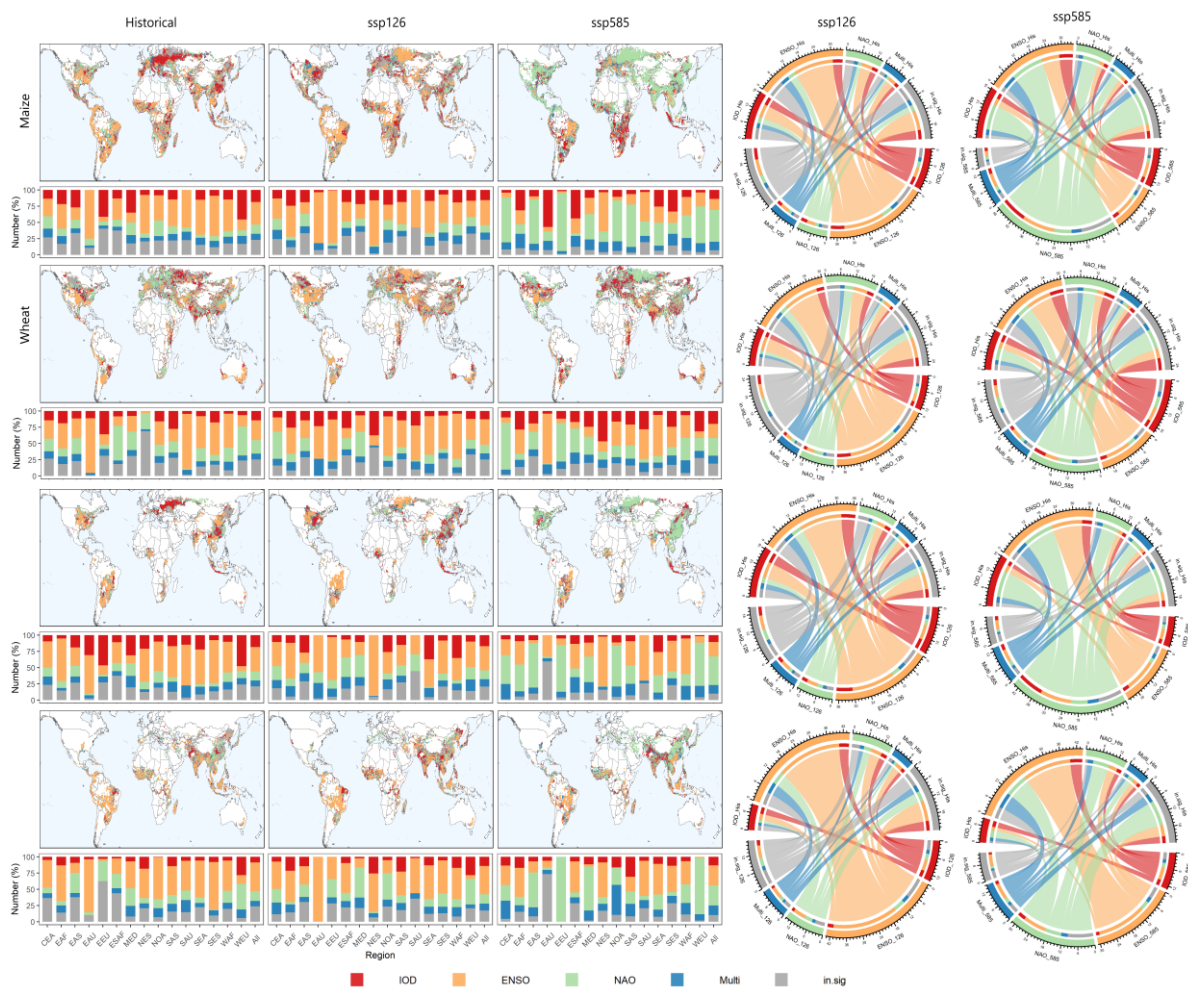
Fig. S2 Dominant climate oscillation indices of global maize, wheat, soybean and rice yield at each grid as identified by the RF model based on IPSL-CM6A-LR



附图 3 基于 MPI-ESM1-2-HR 气候模式，由 RF 模型确定的每个网格的全球玉米和小麦产量的主导大气环流指数

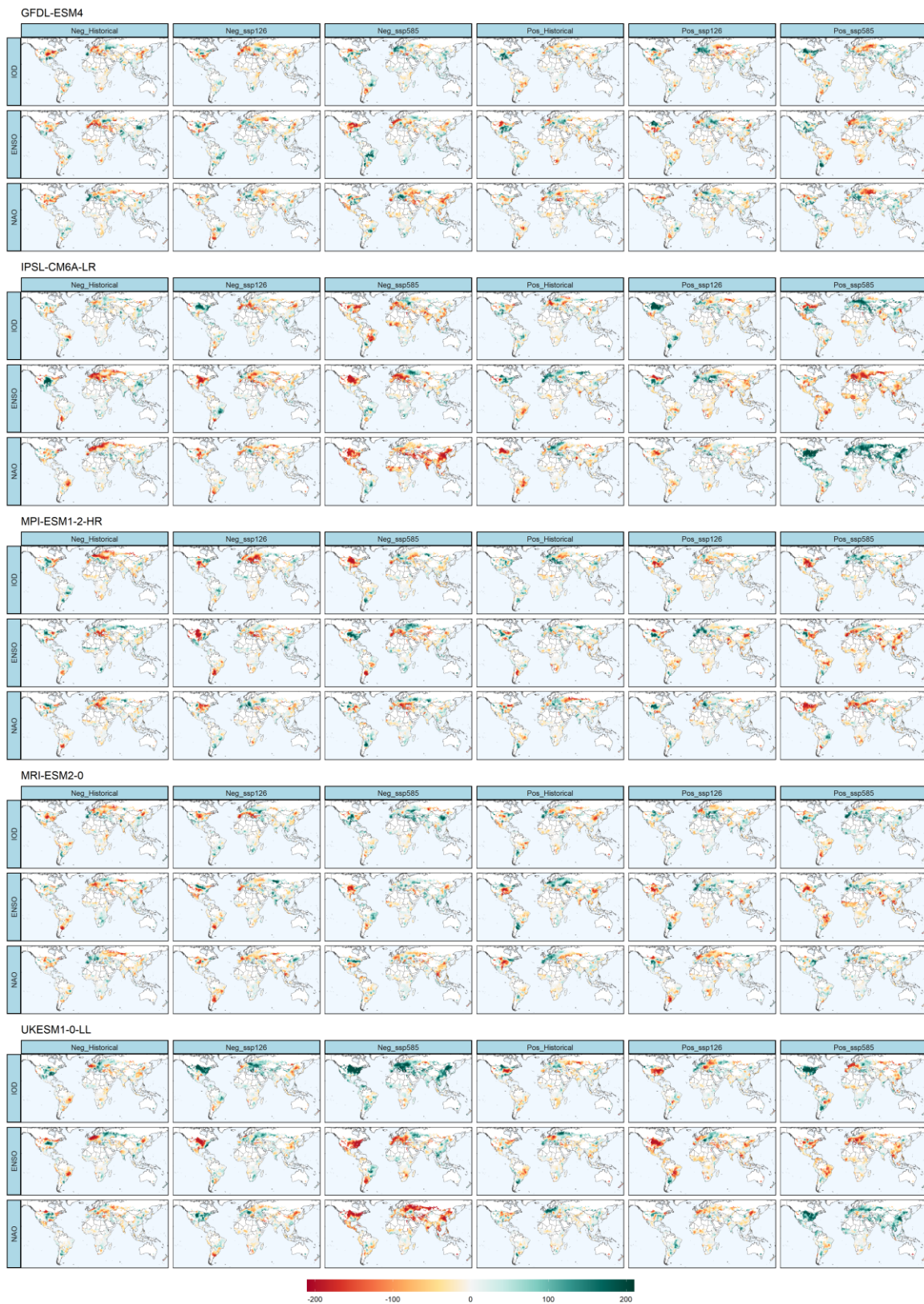
Fig. S3 Dominant climate oscillation indices of global maize, wheat, soybean and rice yield at each grid as identified by the RF model based on MPI-ESM1-2-HR





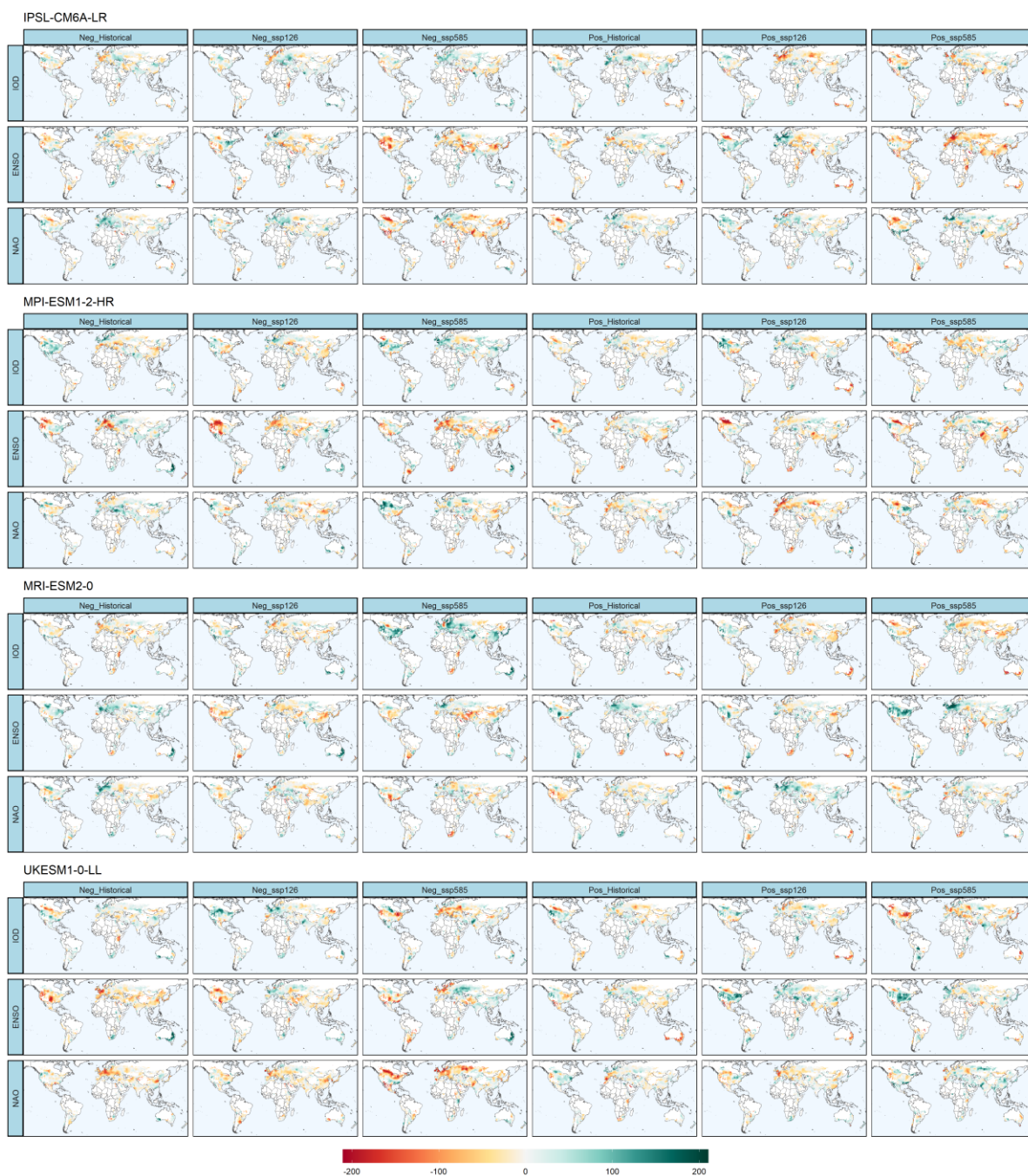
附图 4 基于 UKESM1-0-LL 气候模式，由 RF 模型确定的每个网格的全球玉米和小麦产量的主导大气环流指数

Fig. S4 Dominant climate oscillation indices of global maize, wheat, soybean and rice yield at each grid as identified by the RF model based on UKESM1-0-LL



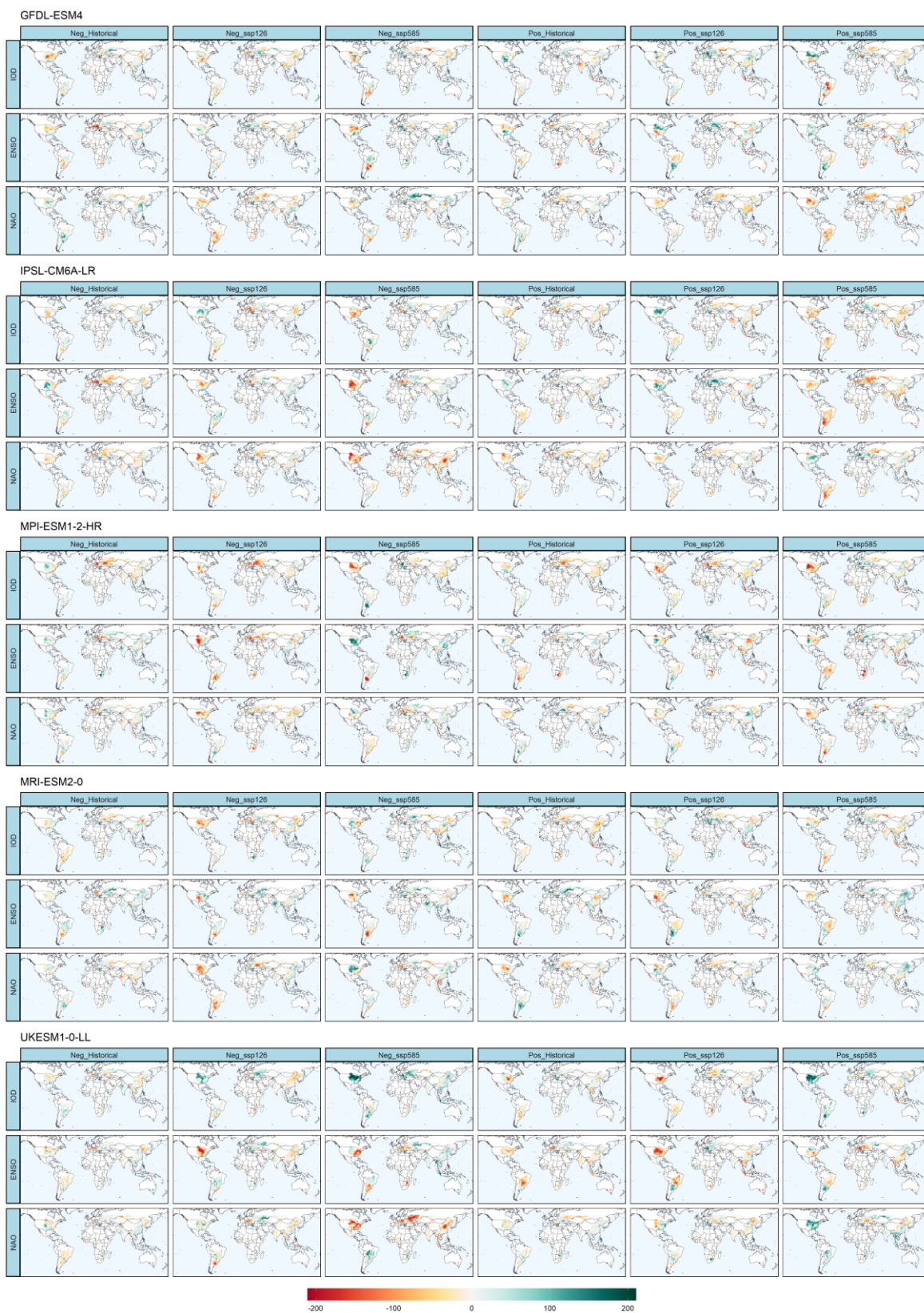
附图 5 五个气候模式下，气候强烈震荡对玉米历史和未来产量的影响

Fig. S5 Maize yield change during strong oscillation phases quantified by partial dependence plot based on 5 GCMs in historical and future under SSP126 and SSP585.



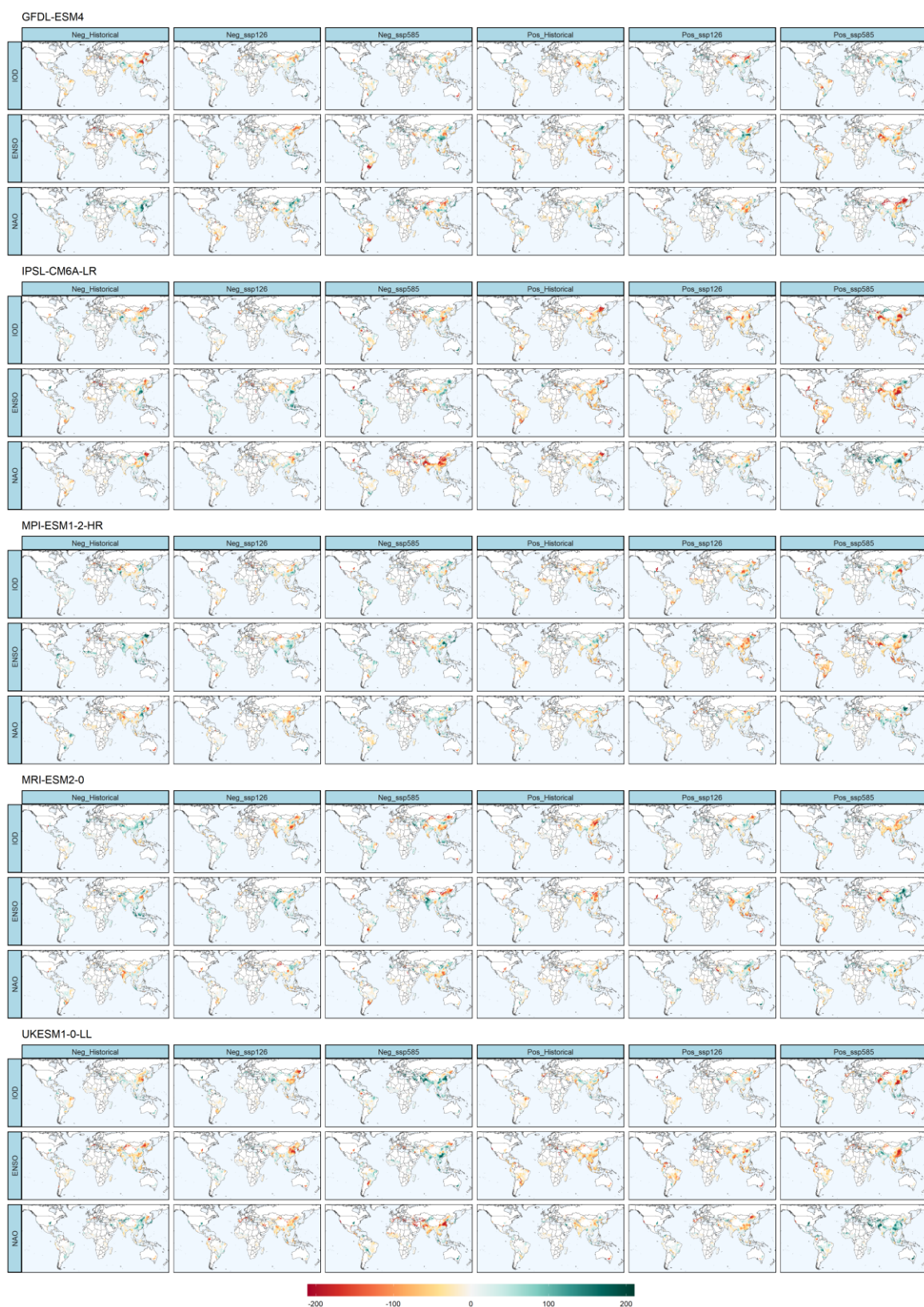
附图 6 余下四个气候模式下，气候强烈震荡对小麦历史和未来产量的影响  
 Fig. S6 Wheat yield change during strong oscillation phases quantified by partial dependence plot based on other 4 GCMs in historical and future under SSP126 and SSP585.





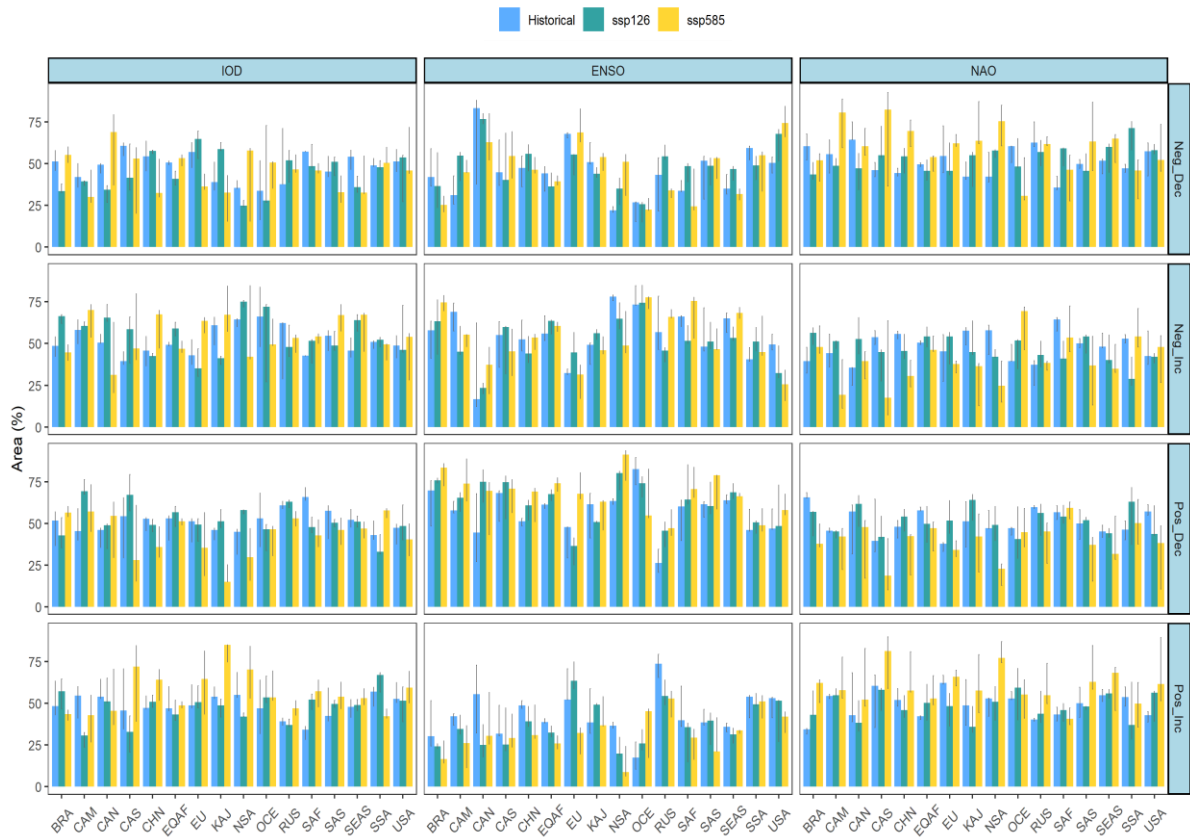
附图 7 五个气候模式下，气候强烈震荡对大豆历史和未来产量的影响

Fig. S7 Soybean yield change during strong oscillation phases quantified by partial dependence plot based on 5 GCMs in historical and future under SSP126 and SSP585.



附图 8 五个气候模式下，气候强烈震荡对水稻历史和未来产量的影响

Fig. S8 Rice yield change during strong oscillation phases quantified by partial dependence plot based on 5 GCMs in historical and future under SSP126 and SSP585.



附图 9 基于历史和未来 5 个 GCMs(SSP126 和 SSP585), 在强振荡阶段玉米产量不同子区域的显著变化的面积。误差条表示 5 个 GCM 的标准误差

Fig. S9 Maize yield significant change areas during strong oscillation phases based on 5 GCMs during the historical and future (SSP126 and SSP585) in different sub-regions. The error bars represent the standard errors for 5 GCMs



附图 10 基于历史和未来 5 个 GCMs(SSP126 和 SSP585)，在强振荡阶段小麦产量不同子区域的显著变化的面积。误差条表示 5 个 GCM 的标准误差

Fig. 10 Wheat yield significant change areas during strong oscillation phases based on 5 GCMs during the historical and future (SSP126 and SSP585) in different sub-regions. The error bars represent the standard errors for 5 GCMs



附图 11 基于历史和未来 5 个 GCMs(SSP126 和 SSP585), 在强振荡阶段大豆产量不同子区域的显著变化的面积。误差条表示 5 个 GCM 的标准误差

Fig. 11 Soybean yield significant change areas during strong oscillation phases based on 5 GCMs during the historical and future (SSP126 and SSP585) in different sub-regions. The error bars represent the standard errors for 5 GCMs

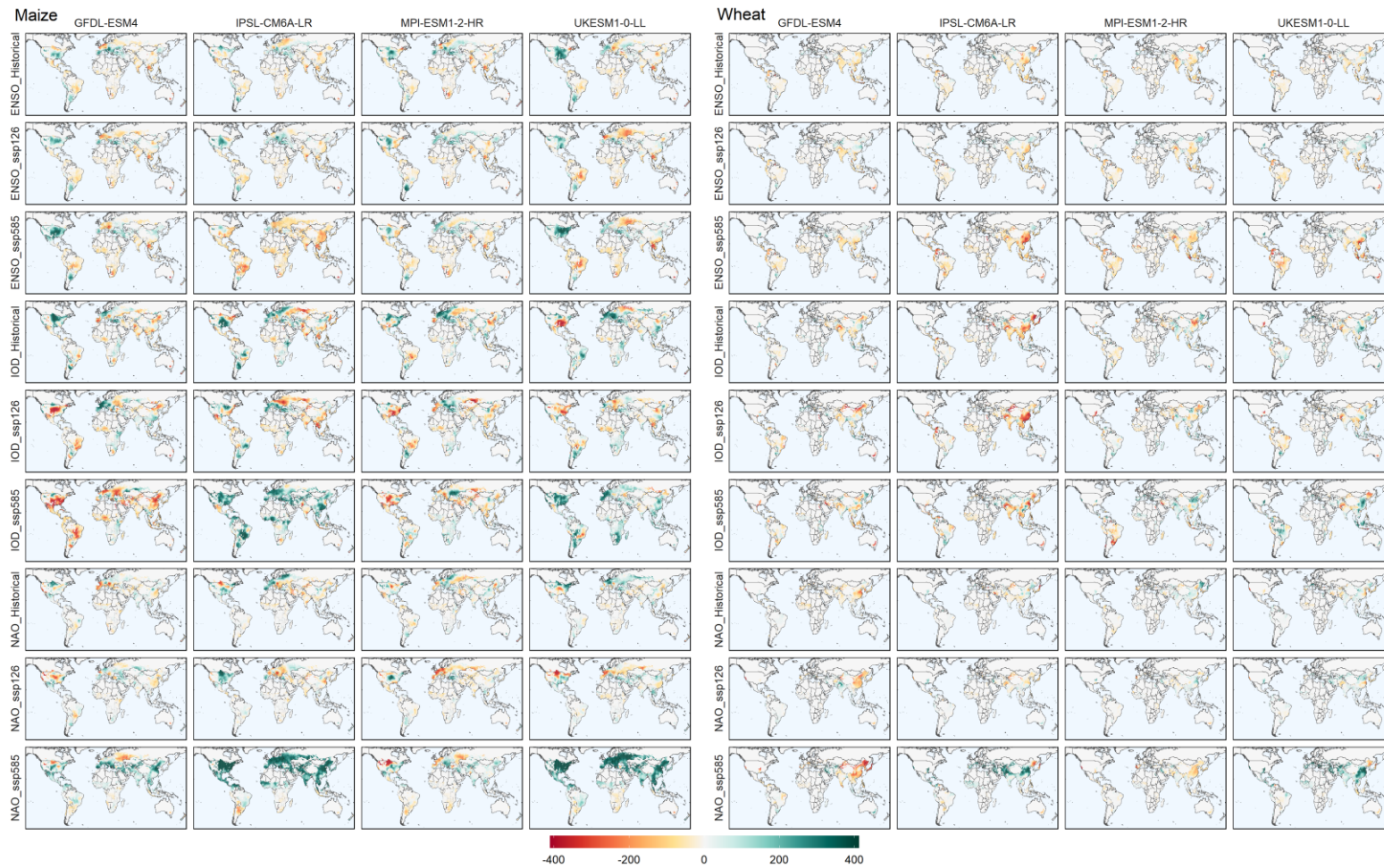




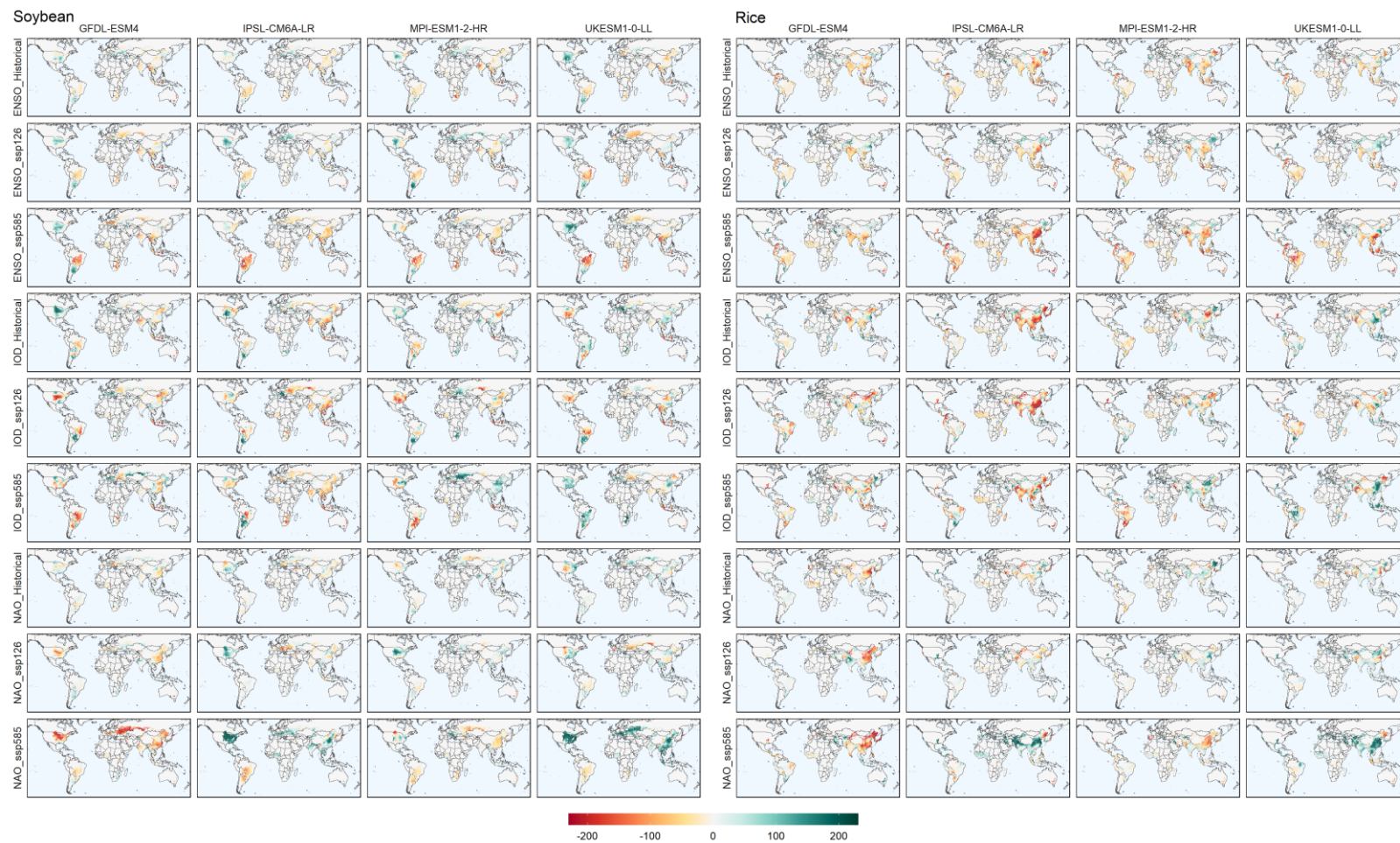
附图 12 基于历史和未来 5 个 GCMs(SSP126 和 SSP585), 在强振荡阶段水稻产量不同子区域的显著变化的面积。误差条表示 5 个 GCM 的标准误差

Fig. 12 Rice yield significant change areas during strong oscillation phases based on 5 GCMs during the historical and future (SSP126 and SSP585) in different sub-regions. The error bars represent the standard errors for 5 GCMs





附图 13 历史(1900-1999)和未来(SSP585, 2000-2099)玉米和小麦产量对 IOD、ENSO 和 NAO 的敏感性。敏感性值由基于其余 4 个 GCM 的 DLM 计算  
Fig. S13 Maize and wheat yield sensitivity to IOD, ENSO, and NAO during historical (1900-1999) and future (SSP585, 2000-2099). The sensitivity values are calculated by DLM based on other 4 GCMs



附图 14 历史(1900-1999)和未来(SSP585, 2000-2099)大豆和水稻产量对 IOD、ENSO 和 NAO 的敏感性。敏感性值由基于其余 4 个 GCM 的 DLM 计算  
 Fig. S14 Soybean and Rice yield sensitivity to IOD, ENSO, and NAO during historical (1900-1999) and future (SSP585, 2000-2099). The sensitivity values are calculated by DLM based on other 4 GCMs

## 致谢

博士四年转眼间已到了尾声，这是我遇到挑战最多也是成长最快的四年，对我有特殊意义。博士生涯如同与另一个自己下棋，只有不断超越自己才能克服困难并有所收获。然而每一次成长都离不开老师、同学以及家人的帮助与陪伴。

非常感谢于强老师在我学术道路上的引导和支持。于老师展现出了卓越的格局和深厚的学术造诣，对问题的把握和分析能力非常出色。在与于老师的交流中，我深深感受到了他对科研的热忱和追求卓越的态度，这激发了我对科学的无限热爱。于强老师营造了一种宽松而充满活力的学术氛围，这让我敢于尝试自己感兴趣的研究方向，并从中发现了一些有趣的科学问题。他鼓励学生独立思考、勇于质疑并解决问题，这培养了我们的创新思维和科研能力。于老师在学术讨论中的启发和指导让我不断成长，并为我打开了更广阔的研究视野。在与于老师的交流中，我学到了如何从一个更宏观的角度看待问题，如何从一个更系统的角度思考科研工作。他的格局之大和对科研问题的把握能力给我留下了深刻的印象。于老师始终鼓励我们挑战自己，勇于探索未知领域，这为我提供了不断超越自我的机会，让我受益匪浅。

感谢杨贵军老师在数据上给予的支持和鼓励，让我的科研得以顺利开展。杨老师的专业知识和经验对我起到了至关重要的作用。同时，我还要特别感谢刘德立老师在数据分析方面给予我的指导和启发。刘老师以高尚的师德和不知疲倦的教学态度对我影响深远，他朴实无华、平易近人的人格魅力让我倍感亲近和敬佩。此外，感谢李毅老师对我的论文给予的指导和帮助。李老师严谨科学的态度、专注认真的精神以及对学术追求的不懈努力对我产生了深刻的影响。感谢王斌和冯璞玉师兄在小论文的设计与写作方面给予的帮助和支持。他们对我的研究项目提出了宝贵的建议和意见，使我学会了如何从科学问题的角度出发去完成一篇论文。也感谢冯浩老师、何建强老师、岳超老师、吕肖良老师、王自奎老师对论文提出的宝贵意见。正是有了以上老师和师兄们的悉心指导和鼓励，我才能克服困难，取得了科研道路上的进步。他们的影响将激励我继续追求知识，不断成长和发展。

感谢姚宁师兄、张亚捷师兄、黄明霞师姐、史利洁师姐、庄伟师兄、刘淮桥师兄、赵杰师兄、吝海霞师姐、何杰师兄、梁楚涛、张艳师姐、金朝师兄、成功师兄在我学习中给予的热心帮助；感谢同级王亚凯、戎亮秉、贾倩兰、薛金儒在学习和生活上提供的无私热情的帮助；感谢施宇、郭晨辉、向柯宇、柏延文、任凯、田琪、谭姣姣、姚林佳、刘一科、杨晶晶、闫雨杏、王欢欢、牛文浩师弟师妹的在生活学习中，给予我无私热情的帮助。感谢女友何沁锶的陪伴使我渡过了最难熬的一段时光。我们相互鼓励，共同进步，使生活变得丰富多彩。

感谢学习生活七年的母校，西北农林科技大学。母校以其优良的学习风气和严谨的科研氛围教会了我如何追求学问。校训“诚朴勇毅”使我明白了承前启后的重要性。宽广包容的情怀和充实浪漫的校园生活也拓宽了我的眼界。

感谢父母，在生活和精神上给予我帮助和爱护。他们是我求学道路上最坚实的支撑，尤其是在我追求学业的过程中，他们给予了我巨大的支持和精神上的鼓励。在儿是多年的求学的道路上，父母一直是最坚强的后盾。他们默默地付出，无私地支持，用自己的爱和关怀为我打造了前行的道路。

最后感谢参加论文评阅和答辩的所有老师，感谢您们的辛苦付出。论文完成之际，再一次向帮助过我的老师、同学、朋友和亲人们表示衷心的感谢！

李林超

2023年5月

## 作者简介

李林超，男，汉族，内蒙古呼和浩特人，本科毕业于内蒙古农业大学水利与土木建筑工程学院，研究生阶段师从西北农林科技大学水建学院李毅教授，从事农业水土资源利用与保护的研究。博士阶段师从西北农林科技大学资环学院于强教授，从事气候变化对农业影响的研究。

### 一、主要学习经历

2012.09-2016.07，内蒙古农业大学，农业水利工程专业，攻读学士学位；

2016.09-2019.07，西北农林科技大学水利工程专业，攻读硕士学位。

2019.09-2023.07，西北农林科技大学土地资源与空间信息技术专业，攻读博士学位

### 二、发表论文

[1] **Linchao Li**, Bin wang, Puyu Feng, Huanhuan Wang, Qinsi He, et al. (2021). Crop yield forecasting and associated optimum lead time analysis based on multi-source environmental data across China. *Agricultural and Forest Meteorology*, 308-309. doi:10.1016/j.agrformet.2021.108558.

[2] **Linchao Li**, Bin wang, Puyu Feng, De Li Liu, Qinsi He, et al. (2022). Developing machine learning models with multi-source environmental data to predict wheat yield in China. *Computers and Electronics in Agriculture*, 194. doi:10.1016/j.compag.2022.106790

[3] **Linchao Li**, Ning Yao, Yi Li, De Li Liu, Bin Wang et al. (2019). Future projections of extreme temperature events in different sub-regions of China. *Atmospheric Research*, 217, 150-164. doi:10.1016/j.atmosres.2018.10.019

[4] **Linchao Li**, Ning Yao, De Li Liu, Songbai Song, Haixia Lin et al. (2019). Historical and future projected frequency of extreme precipitation indicators using the optimized cumulative distribution functions in China. *Journal of Hydrology*, 579. doi:10.1016/j.jhydrol.2019.124170

[5] **Linchao Li**, Yufeng Zhou, Yi Li, Haixia Lin, De Li Liu et al. (2020). Trends, change points and spatial variability in extreme precipitation events from 1961 to 2017 in China. *Hydrology Research*. doi:10.2166/nh.2020.095

[6] 秦榕, **李林超**\*, 杨霁, 杨华, 杨艳玲, 何亚平. (2018) 新疆地区蒸发皿蒸发量变化及基于小波的周期分析. *干旱区地理*. 41(05),954-962.

### 三、学术交流

[1] 中国农业大学农业系统模拟与农业气象暑期学校, 2019

[2] 国际农业模型比较与改进及气候智慧型农业东亚研讨会暨第六届农业模型开发和应用研讨会, 2019

[3] 线上参加第二届中国农业大学农业系统模拟与农业气象暑期学校, 2020

[4] 澳大利亚新南威尔士州初级产业部联合培养, 2022-2023