



西北农林科技大学
NORTHWEST A&F UNIVERSITY

博士学位论文

基于 GIS 与地理空间机器学习的澳大利亚
小麦产量空间预测和异质性研究

学科专业：土地资源与空间信息技术

研究方向：农业 GIS 与地理空间分析

论文作者：任凯

指导教师：于强 研究员

联合指导教师：宋泳泽 副教授

论文提交时间：2025 年 9 月

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

Spatial Prediction and Heterogeneity of Australian Wheat Production Based on GIS and Geospatial Machine Learning

Major: Land Resources and Spatial Information Technology

Research Field: Agricultural GIS and Geospatial Analysis

Candidate: Kai Ren

Supervisor: Prof. Qiang Yu

Co-Supervisor: Dr. Yonzge Song

Date of Submission: Sept., 2025

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: 2020060321

西北农林科技大学博士学位论文

基于 GIS 与地理空间机器学习的澳大利亚 小麦产量空间预测和异质性研究

论文作者: 任 凯

指导教师: 于强 研究员

指导小组: 宋泳泽 副教授 于强 研究员

答辩委员会:

西北农林科技大学水利与建筑工程学院 何建强 教授(主席)

西北农林科技大学水利与建筑工程学院 李毅 研 究 员

西北农林科技大学水土保持科学与工程学院 吴耕泓 教 授

中国农业大学土地科学与技术学院 冯璞玉 副 教 授

中国科学院大学生态环境研究中心 石浩 副 研 究 员

答辩日期: 2025 年 9 月 3 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：2025年9月5日

导师指导研究生学位论文的承诺

本人承诺：我的研究生任凯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：2025年9月5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内 容授权汇编录入《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: 2025 年 9 月 5 日

导师签名:  时间: 2025 年 9 月 5 日

摘要

随着全球气候变异加剧与农业资源环境约束日益突出，深入揭示作物产量的空间异质性并提升产量预测精度，对保障粮食安全和指导精准农业具有重要意义。以澳大利亚主要粮食产区为对象，本文在充分整合多源遥感、气象和土壤大数据的基础上，面向“空间模式识别—预测精度提升—异质性量化”这一主线展开系统研究。首先，构建融合空间自相关与空间分层异质性的地理空间机器学习框架，结合决策树自动离散化和地理探测器解释力评估，形成基于地理最佳区域的异质性（Geographically Optimal Zones-based Heterogeneity, GOZH）模型，用于识别产量热点-冷点格局及其主导环境因子。其次，提出“第二维度异常”（Second-dimension Outliers, SDO）概念及多尺度特征提取流程，将局部异常的第二维度空间信息显式注入机器学习模型，有效解决传统方法对极端高低产区域响应不足的问题，并有效的提高了空间预测的精度和降低了误差。进而在 SDO 概念的基础上，耦合决策树与地理探测器，提出第二维度异常驱动的异质性量化模型（Second-dimension Outlier-driven Heterogeneity, SOH），实现对单变量及交互变量在不同空间模式上的解释力量化，并应用于小麦与大麦产量数据以检验模型的泛化与迁移能力。通过多套对比实验与交叉验证，系统评价了所提方法在空间分区、预测精度和异质性解释方面的优势，为作物在区域化种植管理和可持续农业决策中提供了理论支撑。本研究的主要结论如下：

（1）小麦产量空间差异与决定因素识别

依托融合空间自相关、空间分层异质性与决策树的 GOZH 模型，对澳大利亚小麦带 2015-16 与 2020-21 产量格局与空间差异性进行分析。模型将研究区自动划分为四个地理最优区域，区际差异解释率显著提高；气候（温度、降水）、地形（海拔、坡度）、土壤（有机碳、质地）和植被环境等要素被确定为影响小麦产量空间差异性的决定因素，这为科学的分区化种植管理提供了科学依据。

（2）第二维度异常（SDO）提高小麦产量预测精度

本研究开发了基于样本点以外的空间异常信息的第二维度异常模型（SDO），引入不同缓冲区下的 SDO 变量后，支持向量机模型对小麦产量的五折交叉验证 R^2 提升达到了 33.9%，RMSE 与 MAE 均下降约 20%。结果表明，将局部空间异常信息纳入机器学习模型，能够有效弥补传统机器学习对极端高低产区域响应不足的问题，显著增强了空间预测的准确性与稳健性。

（3）提出 SOH 模型并应用于小麦和大麦的空间异质性量化和解释的案例中

基于第二维度异常（SDO）概念，结合决策树与地理探测器构建了第二维度异常

驱动的异质性量化模型 (SOH)。通过澳大利亚小麦产量案例,模型在结合原始解释变量 (Vars) 后对小麦产量空间分布模式实现了 0.688 的整体解释力,较仅使用原始变量提升 8.3%,表明局部异常信息 (空间模式变量, SOPs) 与环境因子的耦合显著增强了空间异质性解释能力。进一步将 SOH 模型应用于 2023-24 年澳大利亚大麦产量数据,整体解释力达到 0.732,较原始变量基准模型提高 6.1%。该跨作物验证显示 SOH 框架具有良好的迁移性与泛化潜力,可扩展用于多粮种精细化管理与区域可持续农业决策支持。

关键词: 农作物产量; 地理空间分析; 产量预测; 机器学习; 空间异质性

ABSTRACT

With global climate variability intensifying and constraints on agricultural resources and the environment becoming increasingly severe, accurately revealing the spatial heterogeneity of crop yield and improving yield-prediction accuracy are essential for ensuring food security and advancing precision agriculture. Focusing on Australia's principal grain-producing regions, this study integrates multi-source remote-sensing, meteorological and soil big data and conducts a systematic investigation along the main line of "spatial-pattern identification - prediction-accuracy enhancement - heterogeneity quantification." First, a geospatial machine-learning framework that combines spatial autocorrelation with spatial stratified heterogeneity is constructed; together with automated decision-tree discretisation and Geographical Detector assessment, this framework forms the geographically optimal zones-based heterogeneity (GOZH) model for identifying yield hot-cold spots and their dominant environmental factors. Second, the concept of second-dimension outliers (SDO) and a multi-scale feature-extraction procedure are proposed, explicitly incorporating local second-dimension spatial anomaly information into machine-learning models to address the limited response of traditional methods to extreme high- and low-yield areas, thereby markedly improving prediction accuracy and reducing error. Building upon the SDO concept, a second-dimension outlier-driven heterogeneity (SOH) model is developed by coupling decision trees with the Geographical Detector, enabling the quantitative assessment of single and interacting variables across different spatial patterns. The model is applied to wheat and barley yield data to examine its generalisation and transferability. Through multiple comparative experiments and cross-validation, the proposed approach is comprehensively evaluated with respect to spatial partitioning, prediction accuracy and heterogeneity explanation, providing theoretical support for regional crop management and sustainable-agriculture decision-making. The main findings are as follows:

1. Identification of spatial differences and determinants of wheat yield

Using the GOZH model, spatial patterns and heterogeneity of wheat yield in the Australian wheat belt for 2015–16 and 2020–21 were analysed. The model automatically delineated four geographically optimal zones, substantially increasing inter-zone explanatory power; climate (temperature, precipitation), topography (elevation, slope), soil (organic carbon, texture) and vegetation factors were identified as key determinants of spatial yield differences, providing a scientific basis for zone-specific management.

2. Improving wheat-yield prediction accuracy with SDO

A second-dimension outlier model (SDO) based on spatial anomaly information outside the sample points was developed. After introducing SDO variables at different buffer distances, the five-fold cross-validated R^2 of the support-vector-machine model for wheat yield increased by 33.9 %, while RMSE and MAE both decreased by about 20 %. These results show that incorporating local spatial anomalies into machine-learning models effectively compensates for the limited response of conventional methods to extreme yield values, significantly enhancing prediction accuracy and robustness.

3. Development of the SOH model and its application to quantifying and explaining spatial heterogeneity in wheat and barley

Building on the Second-Dimension Outlier (SDO) concept, we integrated decision trees and the Geographical Detector to construct the Second-Dimension Outlier – driven Heterogeneity (SOH) model. In the Australian wheat case study, the SOH model—when combining spatial outlier pattern variables (SOPs) with the original environmental predictors (Vars)—achieved an overall explanatory power (PD) of 0.688, an 8.3% improvement relative to using Vars alone. This indicates that coupling local anomaly information (SOPs) with environmental factors substantially enhances our capacity to quantify and explain spatial heterogeneity in yield patterns. We further applied the SOH framework to 2023–24 Australian barley yield data, where the overall PD reached 0.732, representing a 6.1% gain over the baseline model using only original predictors. This cross-crop validation demonstrates that the SOH framework has good transferability and generalization potential, and can be extended to multi-crop precision management and regional sustainable agriculture decision support.

KEY WORDS: Crop yield; geospatial analysis; yield prediction; machine learning; spatial heterogeneity

目 录

| | |
|--|-----|
| 摘 要 | I |
| ABSTRACT | III |
| 第一章 绪论 | 1 |
| 1.1 研究背景与意义 | 1 |
| 1.2 国内外研究进展 | 3 |
| 1.2.1 气候环境及其他因素对小麦产量的影响 | 3 |
| 1.2.2 小麦产量的空间分布模式 | 4 |
| 1.2.3 空间异质性研究方法 | 5 |
| 1.2.4 小麦产量的空间异质性分析 | 5 |
| 1.2.5 基于机器学习的小麦产量的空间预测 | 7 |
| 1.3 拟解决的科学问题 | 8 |
| 1.4 研究内容与技术路线 | 9 |
| 1.4.1 研究内容 | 9 |
| 1.4.2 技术路线 | 10 |
| 第二章 研究数据与方法 | 12 |
| 2.1 研究区与数据总体概况 | 12 |
| 2.2 研究方法 | 13 |
| 2.2.1 空间自相关分析 | 13 |
| 2.2.2 决策树方法 | 13 |
| 2.2.3 随机森林 | 14 |
| 2.2.4 支持向量机 | 15 |
| 2.2.5 空间分层异质性分析 | 15 |
| 2.2.6 用于空间预测的第二维度异常方法 | 16 |
| 2.2.7 第二维度异常值驱动的异质性方法 | 16 |
| 2.3 小结 | 16 |
| 第三章 识别澳大利亚小麦产量的决定性因素及其空间差异性 | 18 |
| 3.1 概述 | 18 |
| 3.2 研究步骤 | 20 |
| 3.3 材料与方法 | 20 |
| 3.3.1 数据来源及预处理 | 20 |

| | |
|-------------------------------------|-----------|
| 3.3.2 莫兰指数和局部空间自相关分析 | 26 |
| 3.3.3 传统的地理探测器模型 | 27 |
| 3.3.4 地理最优区域异质性 (GOZH) 模型 | 28 |
| 3.3.5 识别变量之间的交互影响 | 30 |
| 3.3.6 模型验证与评估方法 | 31 |
| 3.4 结果分析 | 31 |
| 3.4.1 澳大利亚小麦产量的空间分布模式分析 | 31 |
| 3.4.2 单个变量对小麦产量分布的贡献度 | 34 |
| 3.4.3 识别小麦产量的地理最优区域 | 36 |
| 3.4.4 确定变量交互作用的影响 | 37 |
| 3.4.5 模型验证与评估 | 39 |
| 3.5 讨论 | 41 |
| 3.6 小结 | 43 |
| 第四章 第二维度异常方法的构建及应用 | 44 |
| 4.1 概述 | 44 |
| 4.2 第二维度空间异常方法 | 45 |
| 4.2.1 第二维度空间异常的概念 | 45 |
| 4.2.2 SDO 模型 | 45 |
| 4.3 材料与方法 | 48 |
| 4.3.1 数据来源及预处理 | 48 |
| 4.3.2 基于 SDO 模型的模拟数据 | 50 |
| 4.3.3 随机森林回归方法 (RF) | 51 |
| 4.3.4 支持向量机回归方法 (SVM) | 52 |
| 4.3.5 SDO 模型预测小麦产量的实验设计 | 54 |
| 4.3.6 SDO 模型敏感性分析 | 56 |
| 4.4 结果分析 | 57 |
| 4.4.1 SDO 模型模拟数据分析 | 57 |
| 4.4.2 小麦产量的变量选择与相关性分析 | 59 |
| 4.4.3 生成 SDO 变量和基于 SDO 模型构建机器学习预测模型 | 59 |
| 4.4.4 SDO 模型验证与敏感性分析 | 60 |
| 4.4.5 小麦产量的空间预测与评估 | 62 |
| 4.5 讨论 | 66 |
| 4.6 小结 | 68 |

| | |
|------------------------------------|-----|
| 第五章 基于空间异常的异质性量化解释模型 | 69 |
| 5.1 概述 | 69 |
| 5.2 材料与方法 | 70 |
| 5.2.1 数据及预处理 | 70 |
| 5.2.2 SOH 模型：异常值驱动的异质性量化模型 | 70 |
| 5.3 结果分析 | 73 |
| 5.3.1 空间异常模式（SOPs）的空间分布 | 73 |
| 5.3.2 单个变量及 SOP 的解释力 | 73 |
| 5.3.3 SOPs 间的相互作用机制 | 74 |
| 5.3.4 SOPs 与原始变量的交互作用 | 76 |
| 5.3.5 模型总体验证 | 78 |
| 5.4 扩展分析-量化和解释澳大利亚大麦产量空间分布模式 | 79 |
| 5.4.1 数据收集与变量选择 | 79 |
| 5.4.2 大麦产量的 SOPs 的空间分布 | 80 |
| 5.4.3 单个 SOP 与 Var 的解释力分析 | 82 |
| 5.4.4 SOPs 间的交互作用分析 | 84 |
| 5.4.5 SOPs 与 Vars 的交互作用分析 | 84 |
| 5.4.6 SOH 模型整体的验证与比较 | 88 |
| 5.4.7 大麦案例的结果分析 | 88 |
| 5.5 讨论 | 89 |
| 5.6 小结 | 90 |
| 第六章 结论与展望 | 91 |
| 6.1 主要结论 | 91 |
| 6.2 创新点 | 92 |
| 6.3 研究不足与展望 | 92 |
| 参考文献 | 94 |
| 致 谢 | 108 |
| 个人简历 | 110 |

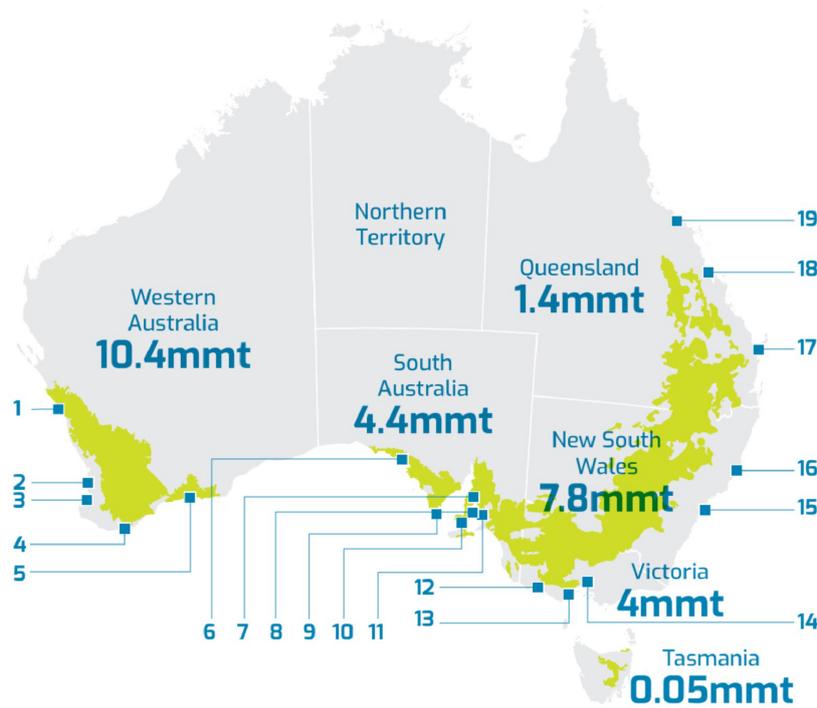
主要符号对照表

| | | |
|--|---------|------------------|
| Geographically Optimal Zones-based Heterogeneity | GOZH | 基于地理最佳区域的异质性模型 |
| Optimal Parameters-based Geographical Detectors | OPGD | 基于最优参数的地理探测器 |
| Second-Dimension Outliers | SDO | 第二维度异常模型 |
| Second-Dimension Outlier-driven Heterogeneity | SOH | 第二维度异常驱动的异质性量化模型 |
| Geographically Optimal Zones | GOZ | 地理最优区域 |
| Geographical Detector | GD | 地理探测器 |
| Power of Determinants | PD | 解释力 |
| Optimal Power of Determinants | OPD | 最优解释力 |
| Outlier Strength Index | OSI | 异常强度指数 |
| Five-fold Cross-Validation | 5-CV | 五折交叉验证 |
| Cross-Validation Root Mean Square Error | CV-RMSE | 交叉验证均方根误差 |
| Spatial Outlier Patterns | SOPs | 空间异常模式 |

第一章 绪论

1.1 研究背景与意义

小麦产量既是衡量农业生产力的重要标尺，也是维护国家粮食安全和全球粮食供给稳定的关键因素。近年来，全球变暖、极端天气事件频发以及水资源短缺等问题叠加，导致小麦生长环境日趋复杂，产量波动显著增大(吴久江 2024)。传统的统计或经验模型难以充分揭示气候要素、土壤条件、地形特征与管理措施等多维因子对小麦产量的综合影响，也难以捕捉局地极端高低产区的空间异质性特征(时运佳等 n.d.)。精准、高效的小麦产量预测不仅能为农户制定播种、施肥与灌溉策略提供参考，还可帮助政府及早评估粮食供需形势，优化储备与调控决策，从而降低粮食安全风险(杨福 2025)。在可持续农业与精准农业快速发展的背景下，迫切需要构建融合遥感、气象、土壤与地理信息技术的新型预测框架，引入对局部空间异常及变量交互作用的刻画手段，以提高产量预测精度、解释产量形成机制，并为区域化管理与适应性调控提供科学支撑。



Source: ABARES five year average to 2023/24 (financial years)

图 1-1 澳大利亚各州的小麦产量

Fig. 1-1 Wheat production in Australian states.

在全球小麦贸易版图中，澳大利亚扮演着“南半球粮仓”的关键角色，其年均小麦

出口量占本国总产量的六成以上，一旦出现产量大幅波动，势必影响亚太乃至全球市场的供需平衡（图 1-1，图 1-2）。然而，澳大利亚小麦带跨越多种气候区——从东部湿润的温带到西部干热的地中海型气候——降水时空分布不稳定，易受厄尔尼诺-拉尼娜等海气耦合现象驱动，加剧了区域间产量差异。同时，土壤质地、西风带移动、热浪频次和耕作制度的差异，使得同一州内甚至相邻农田间的高低产格局呈现显著碎片化特征(童葶 2008)。传统基于州县统计或单一气候指标的预测方法，往往在精细尺度上失灵，难以为农场主提供分区决策，也无法支撑政府对出口配额、仓储物流与价格波动的前瞻性管理(Australian Export Grains Innovation Centre 2025)。

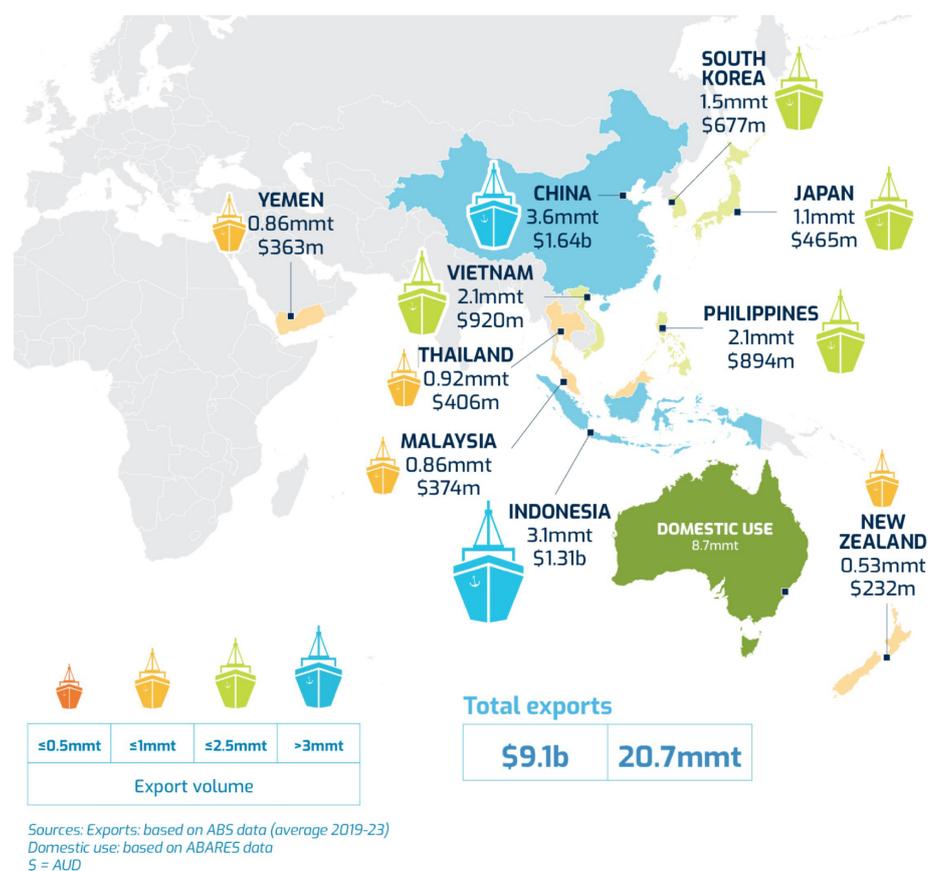


图 1-2 澳大利亚小麦产量的出口数据及其分布

Fig. 1-2 Export data of Australia's wheat production and its distribution.

更具挑战性的是，澳大利亚农业数据呈现“遥感丰富—地面稀疏”的典型特征：卫星影像可高频覆盖，大田实测却分布零散，导致模型构建面临样本稀疏、空间异值突出的问题(张绍华等 2024; 朱行 2016; Henzell 2007; Howden et al. 2010)。如何充分利用遥感—气象—土壤—地形等多源数据，挖掘出局部极端（例如受干旱或突降暴雨影响而异常高低产）的空间模式信息，并将其有效嵌入机器学习模型，是提升预测精度、解释产量形成机理的关键突破口。

基于此背景，本研究以澳大利亚全国尺度小麦产量为切入，开发融合空间自相关分析、空间分层异质性判别与局部异常特征提取的新型地理空间机器学习框架，旨在（一）识别驱动产量空间分异的核心因子及其交互机制；（二）弥补传统模型对极端高低产区响应不足的短板，提高高分辨率产量预测精度；（三）构建具有跨作物迁移能力的异质性量化方法，为澳大利亚乃至其他干旱半干旱地区的多粮种精准管理与气候适应性决策提供可复制的技术路径。

1.2 国内外研究进展

1.2.1 气候环境及其他因素对小麦产量的影响

气候因子始终是调控小麦产量波动的核心驱动力。国际层面，众多实测及模拟研究指出，开花—灌浆期积温决定了潜在穗粒数与千粒重，温度每升高 1°C ，全球小麦平均单产下降3–6%的敏感性已被多模型共识所证实；降水的时空错配被视为澳大利亚、美国西部等半干旱麦区减产的首要风险，而冬春季水分亏缺在北非、西亚麦区又表现为限制播种面积的主要约束(Sharma et al. 2010; Troy et al. 2015; Walker 1989; Wassenaar et al. 1999)。此外，二氧化碳浓度升高对光合同化的“施肥效应”可在一定程度上抵消轻度热胁迫，但在高温、干旱叠加情景下增产效应显著削弱(韦加固等 2025)。针对极端事件的研究也在增多：热浪可缩短灌浆期并诱发蛋白质累积异常，暴雨与雹灾则易造成倒伏及穗发芽，成为近十年风险评估的新焦点。

国内学者长期关注季风区气温上升和降水时序改变对黄淮、东北麦区的影响，发现冬季升温导致越冬级差缩小、返青拔节期提前；华北麦区“暖冬—干春”组合对穗粒数与灌浆速率的抑制最为显著(Hui, Qin, et al. 2019; Xu et al. 2024)。近年来，利用高分辨率再分析资料与遥感指数的研究显示，极端降水频次与SPI（标准化降水指数）对江淮麦区产量的解释力已超越传统季降水总量。

除气候外，土壤、地形及管理措施同样不可忽视。国际上，澳大利亚CSIRO与CIMMYT的试验表明，浅层砂质土的水分保持能力与氮肥损失率直接决定干旱年份的产量底线。坡度与坡向决定了光照与径流分布，在丘陵麦区可造成田块间30%以上的产量差异。播期和品种选择在应对热/旱胁迫中被证实能抵消约15%的潜在损失。国内则更多关注秸秆还田、保护性耕作及深松技术对土壤水分保持与根系分布的改善效果，实践表明适度深松可提高华北平原小麦—玉米轮作系统约8–12%的年均产量(郝展宏等 2025; 刘相荣 2025)。

最新趋势是将多源数据与机器学习方法结合，探索气候—土壤—管理多因子耦合机制(魏煜轩等 2025)。国外如APSIM、DSSAT等过程模型已开始引入再分析气象场与高分辨率土壤网格，国内也出现了融合遥感NDVI、ERA5气象和土壤数字图的随机森

林与深度网络预测框架。然而，大多数模型仍以全局平均或单尺度变量为主，难以捕捉局地极端响应；同时，变量交互及空间异质性解释力的定量刻画不足。

1.2.2 小麦产量的空间分布模式

小麦产量的空间分布模式研究始于 20 世纪初，随着遥感技术与地理信息系统(GIS)的发展，研究方法逐渐从简单的统计分析转向空间自相关、地理加权回归等多尺度空间分析(Basso et al. 2012; Potgieter et al. 2002; Sun et al. 2018)。全球范围内的研究表明，小麦产量受气候、土壤、地形及管理措施等多因素影响，其空间分布呈现出明显的异质性(Han et al. 2024)。

全球尺度的多时序遥感与统计资料表明，小麦产量呈现出明显的带状与块状分布特征：欧洲、中北美与东亚温带区往往形成连续的高产带，而中东、北非及澳大利亚西部则受水分胁迫和土壤贫瘠影响，产量碎片化且波动幅度大(Constantinidou et al. 2016; Heng et al. 2007; Le Gouis et al. 2020; Olmstead and Rhode 2011; Wang et al. 2015; Zampieri et al. 2020)。基于 FAO 与 GAEZ 数据的空间自相关分析显示，高产区之间存在显著的正相关集聚，而干旱、沙质土覆盖区则形成冷热点“镶嵌”格局。

区域尺度研究进一步揭示气候梯度与地形起伏对产量空间模式的塑造作用(Wong and Asseng 2006)。例如，美国大平原由北向南的降水递减导致南部雨养地块产量显著低于灌溉区(Schillerberg et al. 2019)；法国与德国的精细化田块调查发现，坡度超过 12° 或土壤 pH 低于 5.5 的田块常形成低产斑块(Ahrends et al. 2018; Brisson et al. 2010)。澳大利亚研究则指出，“冬季降水—春季高温”组合导致产量在南北向出现条带状反差，局地盐渍土和浅层基岩区又叠加出“点一带”相间的低产团块(Kanso 2010)。

国内在基于县级统计与 1 km 遥感估算结果的研究中发现，华北平原受灌溉网络密度控制，高产区呈规则块状分布；黄淮丘陵与川西盆地因地形破碎，产量分布呈细碎斑块(Chen et al. 2017; Sun et al. 2018)。利用地理探测器的实证显示，热量带、土壤质地和耕层厚度的交互作用能解释黄淮海区 40% 以上的空间变差。

从方法演进看，早期研究多依赖克里金或 IDW 等插值技术，难以捕捉突变边界；随后 Moran's I、LISA 和地理加权回归相继用于揭示局地集聚与参数空间漂移，但对极端值响应偏弱。近年来兴起的随机森林、梯度提升树与深度网络虽显著提升了预测精度，却普遍将空间位置作为普通自变量处理，缺乏对局部异常与分层结构的显式描述。综上，如何在高精度预测的同时定量刻画多尺度空间异质性，是当前关于小麦产量空间分布模式研究的主要瓶颈，也是本文后续方法创新的重要出发点。

1.2.3 空间异质性研究方法

空间异质性描述同一现象在不同地理位置呈现的差异性与不均一性，其量化方法经历了由经典地统计学向机器学习-统计耦合的演进(张幸福等 n.d.; Fletcher 2007; Kolasala and Rollo 1991)。早期研究以变差函数与克里金为代表，通过协方差结构估计对空间连续面进行无偏插值，但要求平稳性且对多源高维数据支持有限(Jalali et al. 2017; Petitti 2001)。随后，基于空间权重矩阵的 Moran's I、Getis-Ord G^* 与 LISA 成为检测集聚与离散的常用手段(Anselin 2019; Peeters et al. 2015; Thompson et al. 2018)；地理加权回归 (GWR) 及其多尺度扩展 (MGWR) 进一步允许局地参数随空间位置变化，但面对强非线性和共线关系表现受限(刘文毅等 2025; Fotheringham et al. 2017; Li et al. 2019)。

近年来，空间计量经济模型 (SLM、SEM、SDM) 在社会经济领域得到广泛应用，能显式分离空间滞后与误差，但对变量离散化与阈值设定较为敏感(Ale-Kajbaf et al. 2024; Jeong et al. 2022; Sardà-Palomera et al. 2012)。地理探测器 (GeoDetector) 通过方差分解直接量化分层异质性，无需线性假设，已被用于气候-地理驱动因子的解释力评估(Hu et al. 2020)。机器学习的兴起推动了随机森林、梯度提升树及深度网络对复杂非线性关系的建模，但传统做法通常只将坐标或邻域均值作为普通特征，缺少对局地异常和多尺度结构的显式表达(Georganos et al. 2021; Xie et al. 2021; Zhu, Yang, et al. 2024)。国外学者开始将多层卷积、图卷积网络与空间权重融合，以捕捉邻域拓扑特征；国内研究也尝试引入多级缓冲统计、异常指标和决策树离散化以改进 GeoDetector，提升对跨尺度交互效应的解释能力。

然而，现有方法在捕捉局地极端、量化多尺度交互和兼顾解释性与预测性方面仍存在不足(Song 2022)。单纯依赖空间自相关或局地回归难以处理非线性与高维数据；机器学习模型虽能提高预测精度，但缺乏对空间异质性的系统量化；而 GeoDetector 等方法在处理复杂交互作用时又面临变量选择与尺度设定的挑战(王劲峰，徐成东 2017)。因此，亟需开发融合空间统计判别、机器学习非线性拟合与局地异常特征提取的新型框架，以更全面刻画作物产量等地理现象的空间异质性。

综上，现有方法在捕捉局部极端、量化多尺度交互和兼顾解释性与预测性方面仍存在不足，亟需开发融合空间统计判别、机器学习非线性拟合与局地异常特征提取的新型框架，以更全面刻画作物产量等地理现象的空间异质性。

1.2.4 小麦产量的空间异质性分析

深入理解小麦产量的空间差异性对于科学决策与气候适应性管理策略制定具有重要意义 (Hernandez-Ochoa et al. 2018)。已有研究表明，气候条件、土壤特征和天气模式等差异可显著影响产量与作物品质，特别是在南澳大利亚及地中海国家 (Diacono et al. 2012)。因此，准确识别驱动小麦产量空间差异的因素，不仅有助于优化资源配置，还可

为区域及跨区域小麦生产的协同发展提供定量支撑 (Lv et al. 2023; Zhang and Li 2022)。

目前已有多种方法从不同角度研究小麦产量的空间差异性, 包括统计方法、空间依赖性分析、异质性建模及空间分层异质性方法。例如, 空间计量经济模型常用于评估气候变化与适应性管理对小麦产量空间分布的影响机制 (Min et al. 2014); 克里金插值与回归分析方法则在识别空间数据的差异性、相关性及自相关方面具有较高精度。随着数据量与变量复杂度的增加, 机器学习与异质性建模算法因其高计算效率与多因子适配能力, 在农业空间分析中应用广泛。

近年来, 空间分层异质性模型作为上述方法的进一步发展, 通过基于属性或规则对空间数据进行分层, 并在分层内探讨异质性特征, 实现更细致、更精准的空间分析 (Luo et al. 2022; Zhang, Li, and Song 2024; Zhang, Song, et al. 2024)。地理探测器方法 (Geographical Detector, GD) 是一种先进的空间分层异质性分析方法, 通过变量分层与区域方差评估, 定量识别主导因素对响应变量的解释力 (Wang et al. 2014, 2010), 已广泛应用于农业生产差异性研究 (Chu et al. 2019; Hou et al. 2023; Zeng et al. 2023)。

在 GD 模型基础上, 参数优化地理探测器模型 (Optimal Parameters-based Geographical Detector, OPGD) 进一步发展了空间数据离散化、空间分层断点与尺度选择过程, 通过最优参数组合提升了空间解释变量中的地理信息提取能力与建模效率 (Song et al. 2020)。例如, 该模型已用于分析中国黄水河流域土壤有机质 (SOM) 空间分布特征及其主导因素 (Liu et al. 2023), 并探讨了影响植被净初级生产力 (NPP) 的驱动因子及其交互作用机制 (Zhang et al. 2023)。然而, OPGD 模型在空间离散过程中并未充分刻画响应变量的空间分布特征, 限制了对解释变量与响应变量关系的深入理解。

为弥补上述不足, Luo 等人提出地理最优分区异质性模型 (Geographically Optimal Zones-based Heterogeneity, GOZH), 通过最大化区域间差异、最小化区域内差异, 实现地理最优区域的识别, 并提升多变量交互作用在空间异质性分析中的稳定性 (Luo et al. 2022)。

影响小麦产量空间差异的自然因素可归纳为四大类。第一类是气候因素, 其通过改变单位面积产量、适宜种植区分布与耕地适宜性, 影响地区间生产潜力与趋势 (Fan et al. 2018; Wang et al. 2018); 第二类为地理因素, 其通过与气候、位置、土壤等变量交互作用, 影响生产格局与农业可持续性; 第三类为土壤属性, 包括土壤质量、水分保持能力与养分供给等, 直接影响作物生长与产量分布 (Ajami et al. 2020; Chu et al. 2019; Luo et al. 2005); 第四类为环境因素, 遥感与地理信息技术可有效监测与分析作物生境、气候波动与生物胁迫, 揭示环境与产量间复杂关系 (Hodson and White 2007; Sbahi et al. 2018)。除单变量效应外, 还需关注多变量间的复杂交互作用 (Ajami et al. 2020; Jiu-jiang et al. 2022)。

此外，小麦产量的空间差异还受尺度效应显著影响，包括农田尺度 (Kravchenko et al. 2017; Mao et al. 2021)、流域尺度 (Mao et al. 2021; Richter et al. 1998)、区域尺度 (Jin et al. 2022; Xiong et al. 2008)、行政尺度 (Fu et al. 2021) 及全球尺度 (He et al. 2022; Porwollik et al. 2017)。遥感与地理空间技术的整合应用，不仅能全面刻画环境因素对产量的影响，也为精准农业管理与政策制定提供科学依据。

尽管当前方法已取得诸多进展，但在大尺度、异构性强的背景下，仍面临诸多挑战。以澳大利亚为例，其地域广阔，气候、土壤与环境差异显著，导致小麦产量在空间上呈现显著异质性 (Fletcher et al. 2020; Orton et al. 2018)。为应对这一复杂性，本文引入 GOZH 模型，以弥补单变量低估、多变量高估及空间划分粗糙等问题，实现关键驱动因子的精准识别与空间影响机制解析 (Hu et al. 2025; Luo et al. 2022)。

1.2.5 基于机器学习的小麦产量的空间预测

空间预测旨在基于地理属性的空间特征（如空间依赖性、分布模式、变异性与协变量）对未知区域的值进行精确估计。空间预测在多个领域具有重要意义，包括地球科学、城市信息学、地理社交媒体分析、农业管理、气象灾害预报与公共健康等 (Din and Yamamoto 2024; Jia et al. 2023; Jiang 2018; Lu et al. 2017; Sen 2016; Zhang, Wang, et al. 2024)。随着地球观测与多源感知技术的迅猛发展，海量地理空间数据不断积累，亟需更加高效、精准的预测方法 (Mishra et al. 2017; Tian et al. 2022; Wadoux et al. 2019)。通过动态空间特征提取与实时空间监测，这些方法能够有效挖掘大数据潜力，生成准确预测结果并提供有价值的空间信息 (Vicente-Serrano et al. 2023; Yin et al. 2023)。高效的空間预测技术有助于深化对地理现象的理解，提升决策支持能力，在不同领域实现精准应用 (Ulloa-Espindola and Perez-Albert 2022; Vicente-Serrano et al. 2023)。

现有空间预测方法可划分为以下几类，各具优势：

第一类为非空间模型 (Aspatial Models)，包括线性回归、随机森林等传统统计方法与机器学习算法 (Georganos and Kalogirou 2022; Taheri Shahraini and Sodoudi 2016)。这类方法适用于通用预测任务，但通常忽略空间关系，导致在处理空间数据时预测能力受限。

第二类为空间依赖模型 (Spatial Dependence Models)，如克里金插值、空间贝叶斯层级模型与地理加性模型 (Strandberg et al. 2019; Wu et al. 2024)。这些方法融合空间自相关性与地理特征，能够捕捉局部变异性。

第三类为空间异质性模型 (Spatial Heterogeneity Models)，代表方法为地理加权回归 (GWR) 及其扩展形式 (Harris et al. 2010; Tan et al. 2017)。该方法通过分析变量关系在空间上的变化揭示局部异质性。

第四类为第二维度空间关联模型 (Second-Dimension Spatial Association, SDA)，该

类方法整合样本点以外的空间信息，提取来自解释变量的“第二维度变量”，以增强预测精度并捕捉更深层次的空间模式 (Song 2022)。

第五类为地理相似性模型 (Geographical Similarity Models)，通过测量地理位置间的相似性来推断目标变量，为多样化地理背景提供灵活的预测方案 (Song 2023; Zhu et al. 2018)。上述各类方法在空间预测中分别应对不同挑战，为深入理解空间数据提供了理论与技术基础。

空间离群值是地理数据中常见的现象，通常表现为极端值、偏离值或潜在错误 (Nirel et al. 1998; Simões and Peterson 2018; Tang et al. 2023)。这类异常值可能违反许多空间模型的基本假设，如线性/非线性回归、克里金法与地统计模型等，其建模准确性将受到干扰 (Berke 2001; Kim et al. 2016; Militino et al. 2006)。空间离群值的存在不仅降低预测准确性，还使空间模型解释结果不可靠。然而，异常值本身可能反映极端天气事件、疾病爆发或交通异常等关键现象 (Kim et al. 2024; Sun et al. 2019; Yang et al. 2020; Yujun et al. 2019; Zhang and Yang 2019)。与传统异常值不同，空间离群值是通过与邻近点局部比较而识别的，其识别与合理处理对于提升预测精度至关重要 (Araki et al. 2017; Baba et al. 2022)。如鲁棒性插值技术（如中位数克里金）允许在保留异常信息的同时减少其负面影响，从而增强模型鲁棒性 (Sun et al. 2019)。

现有针对空间预测的异常值识别方法主要包括以下几类：空间交叉离群检测方法基于不同空间点事件的组合分析，采用 cross K 函数与 Delaunay 三角剖分建立点间关系 (Shi et al. 2018)；Moran 谱离群检测 (MSOD) 通过空间特征向量分析遗传变异分布中的异常模式 (Wagner et al. 2017)；贝叶斯神经网络 (BNNs) 通过不确定性量化实现对异常观测点的识别 (Liu et al. 2010)；双向空间点估计利用 Z 分数衡量估计值与真实值之间的偏差识别离群值 (Wei et al. 2004)；而对于分类型空间数据，Pair Correlation Ratio (PCR) 通过计算跨距离的类别共现频率以揭示离群模式 (Liu et al. 2014)。这些多样化方法在不同应用场景中为异常值识别提供了有效工具。

尽管方法不断发展，当前空间预测研究仍存在诸多不足。首先，许多研究未充分考虑空间异常值在预测中的重要性，导致预测偏差、信息损失与空间变异性低估。其次，缺乏针对局部离群强度的量化与对预测贡献的系统性评估。第三，现有模型往往局限于样本点数据，未能充分利用未知区域的空间信息，致使对极端高值或低值的预测能力受限，影响模型的实用性与泛化能力。

1.3 拟解决的科学问题

在全球气候变化和农业资源环境约束日益严峻的背景下，作物产量的空间分布呈现出高度异质性，且传统预测方法在刻画极端高低产区域方面存在显著局限性。围绕

保障粮食安全与推进精准农业发展的现实需求，本文拟聚焦以下关键科学问题展开系统研究：

(1) 作物产量空间异质性形成机制及主导因素识别问题

作物产量的空间分布受多种自然和人为因素耦合影响，其形成机制复杂，异质性显著。现有研究多基于经验性因果分析，缺乏面向空间分层结构和局部集聚特征的系统识别机制。亟需构建融合空间相关性与地理分区特征的定量分析框架，以识别主导环境因子在不同空间单元下的异质性贡献。

(2) 空间预测模型中局部极端区域响应能力不足问题

当前多数作物产量预测模型以全局模式为主，对局部空间异常区域（如高产或低产极值）响应能力较弱，难以兼顾精度与鲁棒性。如何显式引入样本点以外的空间异常信息，弥补机器学习在边缘样本和空间异值识别中的不足，是提升预测精度的关键科学挑战。

(3) 多源环境因子耦合驱动下的空间异质性交互刻画与跨作物泛化问题

复杂地理环境中，作物产量受气候、地形、土壤与植被等多源因子非线性交互与多尺度空间模式差异共同影响。现有方法多依赖采样点处原始变量，缺乏对局部异常格局的量化，难以科学评估变量间交互对空间分布异质性的贡献；同时，大多数模型局限于单一作物或区域，跨作物、跨生态区间的迁移性与泛化能力尚未系统验证。亟需构建一种能够同时 (i) 刻画多源因子间非线性交互及空间异常模式，并 (ii) 在不同作物与区域间保持稳定解释力与预测性能的地理空间建模框架，以支撑多粮种区域管理与可持续农业决策。

1.4 研究内容与技术路线

1.4.1 研究内容

本论文围绕澳大利亚主要粮食产区，以作物产量空间分布特征为核心，系统开展了以下三个层面的研究工作：

1. 空间差异性识别与决定因素分析

构建融合空间自相关、空间分层异质性与决策树的地理空间机器学习框架（GOZH模型），对小麦产量的热点-冷点格局进行识别，并定量解析气候、地形、土壤、植被等多维环境因子及其交互作用对产量空间差异的贡献。

2. 第二维度异常（SDO）驱动的空间预测方法

提出“第二维度异常”概念，设计多尺度缓冲区空间异常提取流程，将局部极值信息显式融入机器学习模型，开发 SDO-SVM 空间预测方法，以突破传统模型在极端高低产区域响应不足的局限并提高整体预测精度。

3. 异质性量化模型构建与跨作物验证

在 SDO 概念基础上，耦合决策树与地理探测器，提出第二维度异常驱动的异质性量化模型 (SOH)，实现单因子、交互因子及不同类别变量在空间模式下的解释力量化；并分别将模型应用于小麦与大麦产量数据，验证其跨作物的适用性与泛化能力。

通过上述三个研究层面，论文形成了“空间格局识别—预测精度提升—异质性量化”的完整技术链，为区域化种植管理与可持续农业决策提供了新的方法体系和实践框架。

1.4.2 技术路线

本研究围绕“作物产量—气候—地形—土壤—植被”多源大数据，构建了由三个模块串联的整体技术路线 (图 1-3)。首先，在识别小麦产量空间差异性 & 决定因素模块中，对澳大利亚小麦带进行数据统一、主成分分析及异常值处理，依次开展莫兰指数与 LISA 的空间自相关模式分析，利用地理最优区域异质性 (GOZH) 模型量化单因子贡献、区际差异及交互作用，最终通过与 OPGD 的对比检验解释力 (PD) 优劣。其次，在第二维度异常空间预测方法模块中，依据相关性筛选变量，设置多级缓冲区生成第二维度异常 (SDO) 变量，构建 SDO-驱动的多算法机器学习模型，结合交叉验证与敏感性分析评估其对非空间模型的精度提升。第三模块第二维度异常驱动的空间异质性模型，基于 SDO 的概念，提取第二维度空间异常模式 (SOPs)，将 SOPs 融入决策树-地理探测器框架，提出 SOH 模型，对小麦产量的单变量、组合变量及交互变量解释力进行层次化量化，并与传统方法作对比验证。接着，将同一框架迁移至大麦数据，完成数据预处理、SOPs 构建与 SOH 量化分析，检验模型在不同作物和生态区间的泛化能力。三个模块依次实现空间差异识别、预测精度提升与异质性量化，形成了可复制、可迁移的地理空间机器学习体系，为多粮种精准管理与区域可持续农业提供了系统性技术支撑。

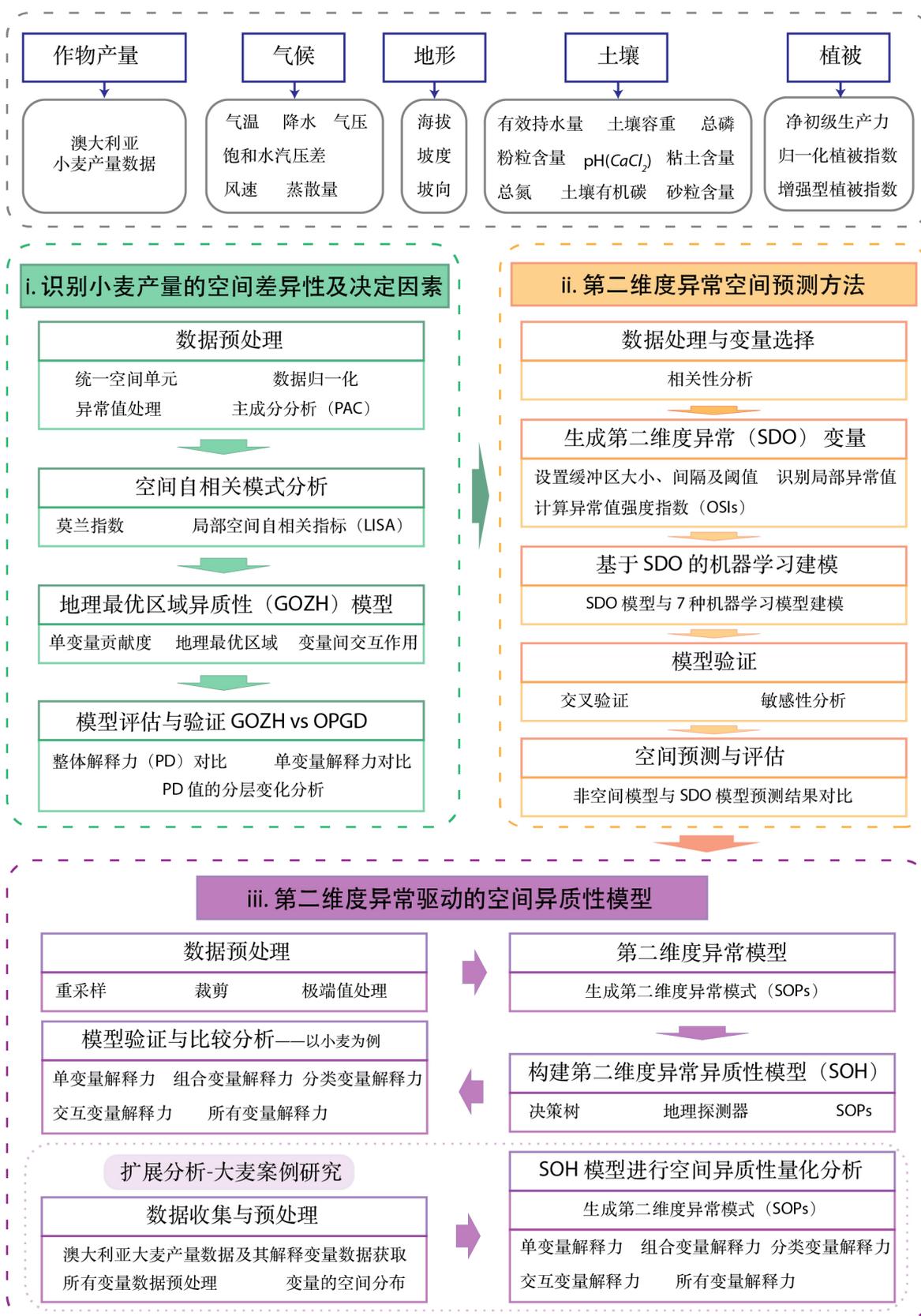


图 1-3 技术路线图

Fig. 1-3 Technical roadmap.

第二章 研究数据与方法

2.1 研究区与数据总体概况

小麦是全球最主要的粮食作物之一，种植面积与种植范围均居世界前列，其产量与分布对全球粮食安全具有重要意义(Food and Agriculture Organization of the United Nations 2019)。澳大利亚是世界上主要的小麦生产国和出口国之一，拥有丰富的种植资源与适宜的自然条件。根据澳大利亚统计局 (Australian Bureau of Statistics, ABS) 数据，2020-2021 财年澳大利亚小麦种植面积达 1,264 万公顷，产量 3,192 万吨，占全球总产量约 5% (Australian Bureau of Agricultural and Resource Economics and Sciences 2025)，产量排名全球第五，仅次于中国、印度、美国和俄罗斯(Australian Export Grains Innovation Centre 2025)。澳大利亚农业和资源经济局 (ABARES) 数据显示，2021 年出口量达 2,500 万吨，占本国总产量约 71%，占全球小麦出口总量的 10% 左右(Department of Agriculture, Fisheries and Forestry 2023)。

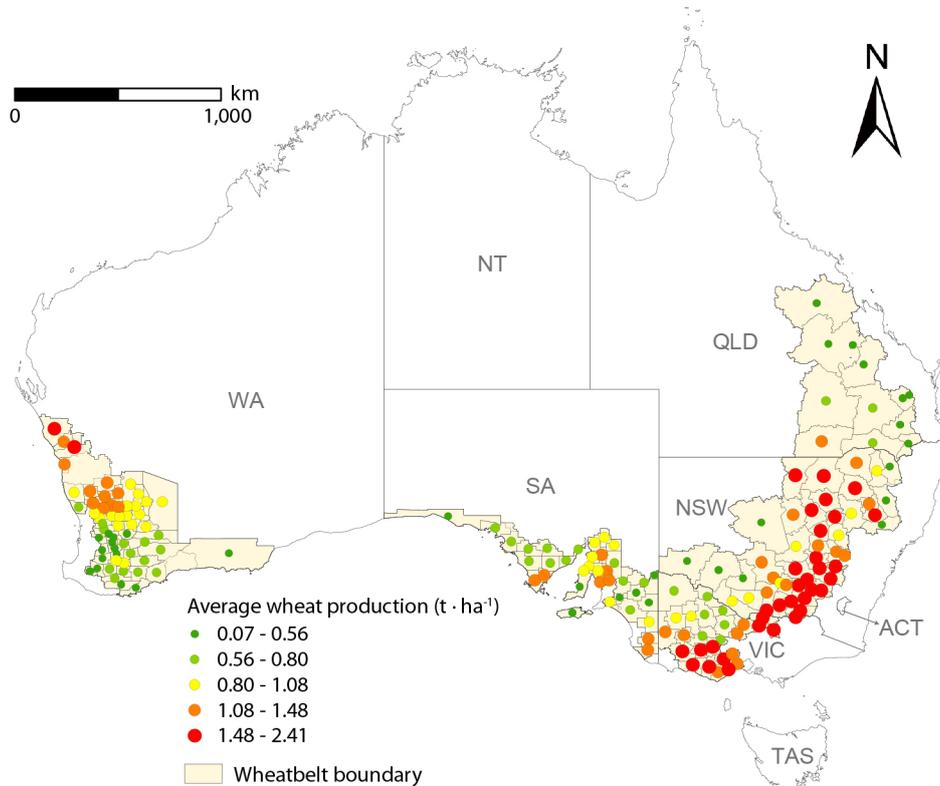


图 2-1 2021 年澳大利亚在 LGA 区域内的平均小麦产量及其空间分布。

Fig. 2-1 Average wheat yield and its spatial distribution within Local Government Areas (LGA) in Australia (2021)

本研究聚焦于澳大利亚的小麦主产区（图 2-1），涵盖新南威尔士州（NSW）、维多利亚州（VIC）、昆士兰州（QLD）、南澳大利亚州（SA）及西澳大利亚州（WA）等。这些区域气候温和、降水适中、日照充足、土壤肥沃，农业生产条件优越，长期保持着较高的小麦产量。研究尺度采用地方政府区域（Local Government Area, LGA）作为空间单元，结合农业普查数据与空间分析方法探讨产量的空间差异及驱动因素。

为分析澳大利亚小麦产量的空间分布特征及其影响因素，研究共构建了四大类解释变量：地理变量、气候变量、土壤属性与环境变量。这些数据主要来源于 Google Earth Engine (GEE) 平台及相关科研机构，如欧洲中期天气预报中心 (ECMWF)、澳大利亚气象局 (BOM)、澳大利亚土壤与景观网格 (SLGA) 等。总体数据框架兼顾了气候、地形、土壤和植被等多方面因素，为后续的空间建模与异质性分析提供支撑。

上述总体介绍为后续章节的详细数据与方法说明奠定了背景和框架。

2.2 研究方法

2.2.1 空间自相关分析

空间自相关分析旨在揭示地理事象在空间上的聚集或离散特征，即判断“近者相似”是否成立(Cliff and Ord 1970; Dormann et al. 2007)。常用的两级指标分别是全局莫兰指数 (Moran's I) 和局部空间自相关指标 (Local Indicators of Spatial Association, LISA)。

Moran's I 用于评估整个研究区的总体空间依赖程度：当指数呈正值时，说明高值与高值、低值与低值倾向聚集；负值表示高低值交错分布，呈离散模式；若接近零，则总体空间随机。研究人员通常借助随机化检验（如蒙特卡罗置换）判断其显著性，以避免将偶然聚集误判为真实结构(Kelejian and Prucha 2001)。

LISA (局部) 则进一步到像元或行政单元层面，识别“热点”(高值-高值)、“冷点”(低值-低值) 以及“离群点”(高值-低值、低值-高值)。LISA 结果常以四象限图或显著性地图可视化，直观呈现局地聚集格局及其显著区域，对精准干预及资源配置尤为重要(Anselin 1995)。

二者互为补充：Moran's I 提供宏观判断，而 LISA 揭示微观空间差异。结合使用，可在定量评估整体空间结构的同时，锁定关键聚集区或异常区，为后续机理解释与决策制定奠定基础(Bivand and Wong 2018)。

2.2.2 决策树方法

决策树 (Decision Tree) 是一种基于递归分割思想的非参数算法，可在多维特征空间内自动生成一套层次化的“如果-那么”规则。其核心流程是：从整张数据集出发，依次选择最能提升区间差异的自变量及划分阈值，将样本划分为同质性更高的左右子集；

随后在每个子集上重复该过程，直至达到预设的停止准则（如最小样本量、最大深度或增益阈值）(Song and Lu 2015)。通过这种逐层细分，决策树能够在不依赖线性假设的前提下，挖掘属性特征与响应变量之间的复杂、非线性关系，并在地理空间场景中形成一组清晰可解释的最优区域(Myles et al. 2004)。

用于空间最优划分时，决策树将地理或环境变量（包括坐标、地形、气候和人文因素等）作为潜在分裂依据，通过“最大化区间差异、最小化区内差异”的准则递归分区。每一个叶节点即代表一块内部同质、外部异质的空间子区，方便后续针对性管理或机理解释(Navada et al. 2011; Suthaharan 2016a)。与传统基于阈值或经验分区的方法相比，决策树具有以下优势：

- **自动化划分**：无需事先指定区域数量或边界，可依据数据本身递归生成多尺度空间分区。
- **高可解释性**：输出的树结构直观展示了变量优先级和阈值，便于追溯影响机制并与领域知识对照。
- **非线性与交互捕捉**：能自然表示非线性关系及多变量交互，无需手动构造复杂项。
- **鲁棒性与灵活性**：对异常值不敏感，可适用于连续型、分类型及混合型特征；通过剪枝和复杂度参数易于控制过拟合。
- **可并行扩展**：若结合装袋或随机特征子集策略，可在大数据环境下并行构建多棵树，实现高效处理。

凭借上述特性，基于决策树的空间最优划分方法广泛应用于生态功能区划、农田管理分区、环境污染分区与风险识别等领域，为精准决策提供了有效的数据驱动工具。

2.2.3 随机森林

随机森林 (Random Forest) 是一种典型的集成学习方法，其核心思想是通过集成大量相互独立、结构多样的决策树来提升模型的预测性能与泛化能力(Belgiu and Drăguț 2016; Rigatti 2017)。它基于装袋 (bagging) 策略，即在原始训练集中利用自助抽样 (bootstrap) 生成多个带有重复样本的子集，并在每棵树的训练过程中对节点划分所使用的特征进行随机筛选，从而有效降低树与树之间的相关性。这种“双重随机性”使得随机森林在面对高维、非线性及噪声较多的数据时仍能保持稳健的表现。其最终预测结果通过对所有决策树的输出进行投票（用于分类任务）或求均值（用于回归任务）得到，从而显著减少单棵树容易出现的过拟合风险。随机森林无需对数据进行复杂的预处理，可直接处理缺失值、异常值及多重共线性问题，并可利用袋外样本 (Out-of-Bag, OOB) 在训练阶段对泛化误差进行无偏估计(Liu et al. 2012; Salman et al. 2024)。此外，该算法能够提供基于置换的变量重要性评价，为解释模型和识别关键影响因素提供了便利。凭借易用性、精度高和适应性强等优势，随机森林已被广泛应用于生态环境监测、作物

产量预测、气候变化评估、遥感影像分类以及公共健康风险评估等多个领域，成为一种兼具理论价值与实践效果的通用型机器学习方法(Rodriguez-Galiano et al. 2015; Segal 2004)。

2.2.4 支持向量机

支持向量机 (Support Vector Machine, SVM) 是一种基于统计学习理论发展的监督式机器学习方法,其核心思想是在特征空间中构造一个能够最大化类别间间隔 (margin) 的最优超平面,从而在保证分类精度的同时提升模型的泛化能力。SVM 遵循“间隔最大化 + 结构风险最小化”的原则,能够在样本量有限且存在噪声的条件下依然获得稳健的分类效果(Hearst et al. 1998; Suthaharan 2016b)。借助核函数 (kernel trick), SVM 可以将输入数据映射到高维甚至无限维的特征空间,从而在该空间中实现线性可分,并在原空间中完成复杂的非线性分类任务。常用的核函数包括线性核、径向基函数 (RBF) 核、多项式核及 Sigmoid 核,不同核函数在隐式映射后的几何结构与决策边界形态上存在显著差异,直接影响模型在不同类型数据集上的表现。凭借较强的理论基础、对高维数据的良好适应性以及对小样本情境的优势, SVM 已广泛应用于模式识别、图像处理、文本分类、生物信息学、环境监测等多个领域(Evgeniou and Pontil 1999; Ma and Guo 2014; Zhang 2012)。

在回归分析中, SVM 被推广为支持向量回归 (Support Vector Regression, SVR), 其核心思想是通过 ϵ -不敏感损失函数 (ϵ -insensitive loss) 在高维特征空间中拟合一个“ ϵ 宽度管”内的最优函数,使预测值与真实值之间的偏差在 ϵ 范围内时不产生损失(Awad and Khanna 2015; Drucker et al. 1996)。这一机制增强了模型对噪声和异常值的鲁棒性,并使得 SVR 在数据存在测量误差或波动时依然保持较高的拟合精度。与传统最小二乘回归不同, SVR 的解由少量支持向量决定,模型复杂度与样本总量无关,因此在高维小样本问题中具有明显优势(Kavitha et al. 2016)。此外, SVR 还可结合不同核函数灵活处理线性与非线性回归问题,在气象预测、经济与金融时间序列分析、农作物产量预估以及工程建模等领域都表现出优异的泛化能力和稳定性(Brereton and Lloyd 2010)。

2.2.5 空间分层异质性分析

空间分层异质性 (Spatial stratified heterogeneity, SSH) 强调“区内同质、区间异质”的地理特征,其基本思路是利用自然断点、分位数或数据驱动算法将研究区划分为若干子区,再比较各子区内部方差与全区总体方差的差异,以评估该分层对整体空间变异的解释程度(Wang et al. 2024)。若区内方差显著低于总体方差,则说明当前分层有效揭示了空间格局;进一步可通过叠加多个解释变量检验它们在空间上的协同或减弱效应,从而探明复杂驱动机制(Bai et al. 2023)。

地理探测器 (Geographical Detector, GD) 是近年提出的一种定量分析 SSH 及其驱动因素的统计工具, 其核心思想建立在“空间一致性假设”之上: 如果某解释变量对响应变量具有显著影响, 则二者在空间上的分布应表现出相似的分层格局(Song et al. 2020)。GD 通过计算 Q 值衡量分层解释力, 并可评估多因子交互作用及划定高风险区。该方法不依赖线性假设、对多重共线性免疫, 已广泛应用于公共健康、生态与环境、农业产量分区等领域, 为揭示空间格局和主导驱动因子提供了可靠依据(Ding et al. 2019; Peng et al. 2019; Qiao et al. 2019)。

2.2.6 用于空间预测的第二维度异常方法

第二维度异常方法是一种新的空间预测方法, 是由本文作者首次提出的。该方法不同于传统的空间预测方法仅依赖样本点自身的空间属性(如自然环境、地理条件、位置信息或人为因素)等信息进行建模, 而是进一步引入了样本点周边邻域的可能导致样本点数据产生变化的空间异常结构信息, 从而构建跨越属性维度和空间结构维度的“第二维度”变量体系(Song 2022)。

该方法的核心思想是分析样本点周围一定范围领域内的空间异常结构信息, 计算这些未知位置的正负异常强度指标, 进而构建一个全新的变量体系。通过对样本点周边的空间异常结构进行分析, 可以更好地理解样本点数据的变化规律, 从而提高预测精度。这个全新的变量体系的数量远远大于原有样本点的数据量, 能够更全面地反映样本点的空间分布特征, 其与原有样本点未知的解释变量数据进行结合从而构建新的第二维度机器学习预测模型, 可以更精确的对未知位置处的目标变量的数值和空间分布进行空间预测。

关于第二维度异常方法的详细介绍和实现过程, 请参见第四章内容。

2.2.7 第二维度异常值驱动的异质性方法

第二维度异常值驱动的异质性模型是本文作者基于上述的第二维度异常方法提出的一种新的空间分布模式及其异质性量化解释的模型。该模型的核心思想是通过分析样本点周边的空间异常结构信息, 通过这些异常值信息计算其对应的异常强度指标, 生成第二维度变量, 其与原有样本点的变量进行组合可以更好的量化和解释样本点的空间异质性和分布模式。

具体的原理、案例分析和实现过程请参见第五章内容。

2.3 小结

本章主要介绍了本研究所用到的数据来源、研究区域、研究方法等内容。通过对澳大利亚小麦主产区的地理、气候、土壤和环境变量进行分析, 结合空间自相关分析、

决策树方法、随机森林方法、支持向量机回归方法、地理探测器等多种方法，为后续的小麦产量预测和空间分布模式分析奠定了基础。同时，提出了第二维度异常方法和第二维度异常值驱动的异质性模型，为空间预测和异质性量化解释提供了新的思路和方法。

第三章 识别澳大利亚小麦产量的决定性因素及其空间差异性

本章基于地理最优区域的异质性模型 (Geographically optimal zones-based heterogeneity model, GOZH), 通过对澳大利亚小麦产量的空间模式进行分析, 识别影响澳大利亚小麦产量的决定性因素及其空间差异性。首先, 通过对澳大利亚小麦产量的全局和局部空间模式进行分析, 确定澳大利亚小麦产量在研究区域上的空间自相关性和空间聚集特征; 其次, 根据单个变量对于小麦产量影响的贡献度, 识别影响澳大利亚小麦产量的决定性因素; 然后, 通过运行 GOZH 模型, 通过决策树等机器学习方法划分小麦产量的地理最优区域; 接着, 分析解释变量之间的交互作用对于小麦产量的空间差异性的影响; 最后, 通过与基于最优参数的地理探测器 (Optimal parameters-based geographical detectors, OPGD) 模型进行对比, 进一步验证了 GOZH 在空间分层异质性分析中的有效性和优势。

3.1 概述

基于地理最优区域的异质性模型 (Geographically optimal zones-based heterogeneity model, GOZH) 是一种通过对解释变量对因变量的影响强度逐级划分地理最优区域的空间分层异质性模型 (Luo et al. 2022)。地理最优区域 (Geographically Optimal Zones) 定义为通过回归决策树对研究区进行逐层划分所得的子区域。经过两次二叉分裂形成四个区域, 每次分裂依据能最大化组内差异减少的变量与阈值自动确定。对各子区域分别进行变量重要性与模型验证, 以识别各区域内响应变量的主要决定因素。GOZH 模型的基本思想是通过地理最优区域的划分, 将研究区域划分为若干个地理最优区域, 然后在每个地理最优区域内分别建立模型, 以此来考察研究区域内不同地理最优区域的异质性特征。本研究中使用 GOZH 模型应用于澳大利亚小麦产量的空间差异性分析中, 其基本框架如图 3-1 所示。

GOZH 模型结合了智能区域划分, 多变量协同交互分析和提升解释力和地理准确性, 克服了传统地理探测模型的不足, 是空间结构与异质性研究的一个更为精细、稳健、高效的方法。其主要有以下优点(Luo et al. 2022):

(1) 最大化组间差异、最小化组内差异

GOZH 在划分地理“最优区域”时, 旨在提高区域间的差异性, 并减少区域内的相似性, 从而更有效地捕捉空间分层异质性。

(2) 避免传统分层中的误差累积

相比传统将多个变量层叠离散化, GOZH 通过逐步空间离散化方法, 降低了因分

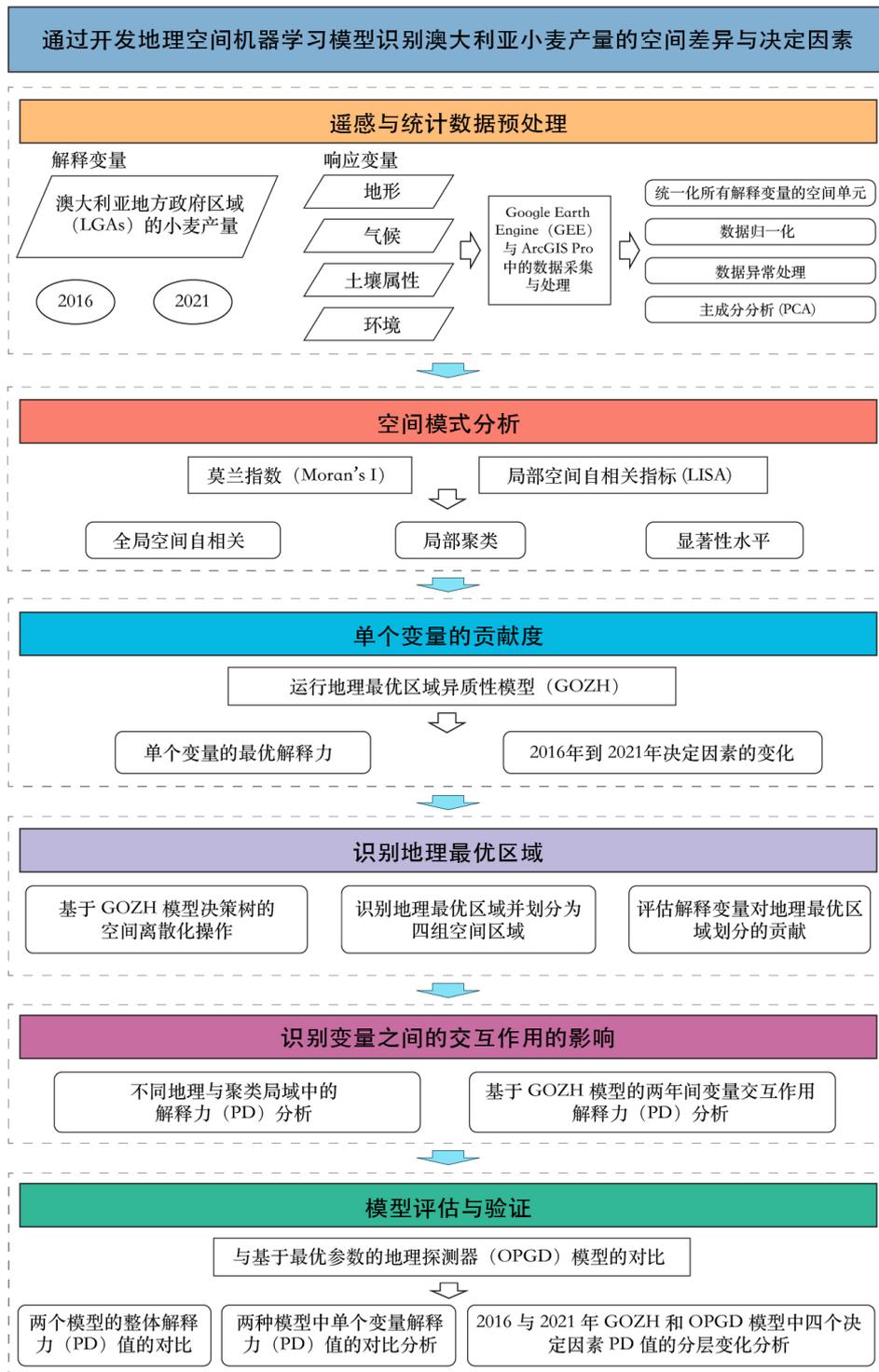


图 3-1 基于地理空间机器学习模型识别澳大利亚小麦产量空间差异及其影响因素的路线图
 Fig. 3-1 Schematic overview of identifying the spatial disparities and determinants of wheat production in Australia based on the geospatial machine learning model.

层参数手动设定而产生的变量高估或低估问题。

(3) 提高交互作用解释的可靠性

传统方法往往只能估计两个或三个解释变量之间的交互作用，且会产生大量碎片化区域；而 GOZH 的区域划分和解释力（Power of Determinants, PD）设计，使得多变量间交互贡献可被更稳定地估算，避免“过度细分”导致的不可靠现象。

（4）兼具全局与局部空间异质性揭示能力

GOZH 同时支持单变量分析和区域划分基础上的多变量交互分析，在识别空间异质性方面更加全面。

（5）相比传统地理探测器模型具备更强解释力

实证中，GOZH 的 PD 值普遍高于传统的地理探测器，尤其是在复杂地理变量和非线性交互变量下，GOZH 在解释率和空间结构识别方面都更具优势。

3.2 研究步骤

本部分研究构建了一种地理空间机器学习模型，用于分析澳大利亚小麦产量的空间差异性，并识别其主导影响因子。该模型综合考虑了空间自相关性、空间分层异质性与决策树分析等方法。图 3-1 展示了本研究的整体研究流程，主要包括六个步骤：第一步，对收集的变量数据进行预处理，以满足模型输入要求；第二步，对响应变量（小麦产量）进行空间自相关分析，评估其热点分布及聚集特征；第三步，评估各个解释变量对小麦产量的单独贡献；第四步，识别澳大利亚小麦产量的地理最优区域，并对其进行区域分组；第五步，分析各解释变量之间的交互作用，计算热点区与各州的解释力（Power of determinants, PD）值；第六步，通过与另一种地理探测器模型进行对比，验证所构建模型的有效性与优势(Ren et al. 2025)。

3.3 材料与方法

3.3.1 数据来源及预处理

3.3.1.1 研究区域与小麦产量数据

小麦是澳大利亚最重要的粮食作物，约占全球年产量的 3.5%，年均总产量约 2,500 万吨 (Australian Export Grains Innovation Centre 2022; Kingwell 2020)。澳大利亚小麦以出口为主，出口量占全国产量的 65–75%，销往全球 50 余个国家，其中西澳大利亚州 (WA) 是最大的小麦出口州 (Australian Export Grains Innovation Centre 2022; Kingwell 2020)。鉴于澳大利亚在小麦生产与出口中的重要地位，本章澳大利亚的小麦产量为研究对象。

本研究选取澳大利亚统计局 (Australian Bureau of Statistics, ABS) 公布的 2015–2016 及 2020–2021 两个农业普查年度的小麦产量数据 (Australian Bureau of Statistics 2017, 2022)。2021 年澳大利亚全国小麦总产量达 3,190 万吨，其中新南威尔士州 (NSW)

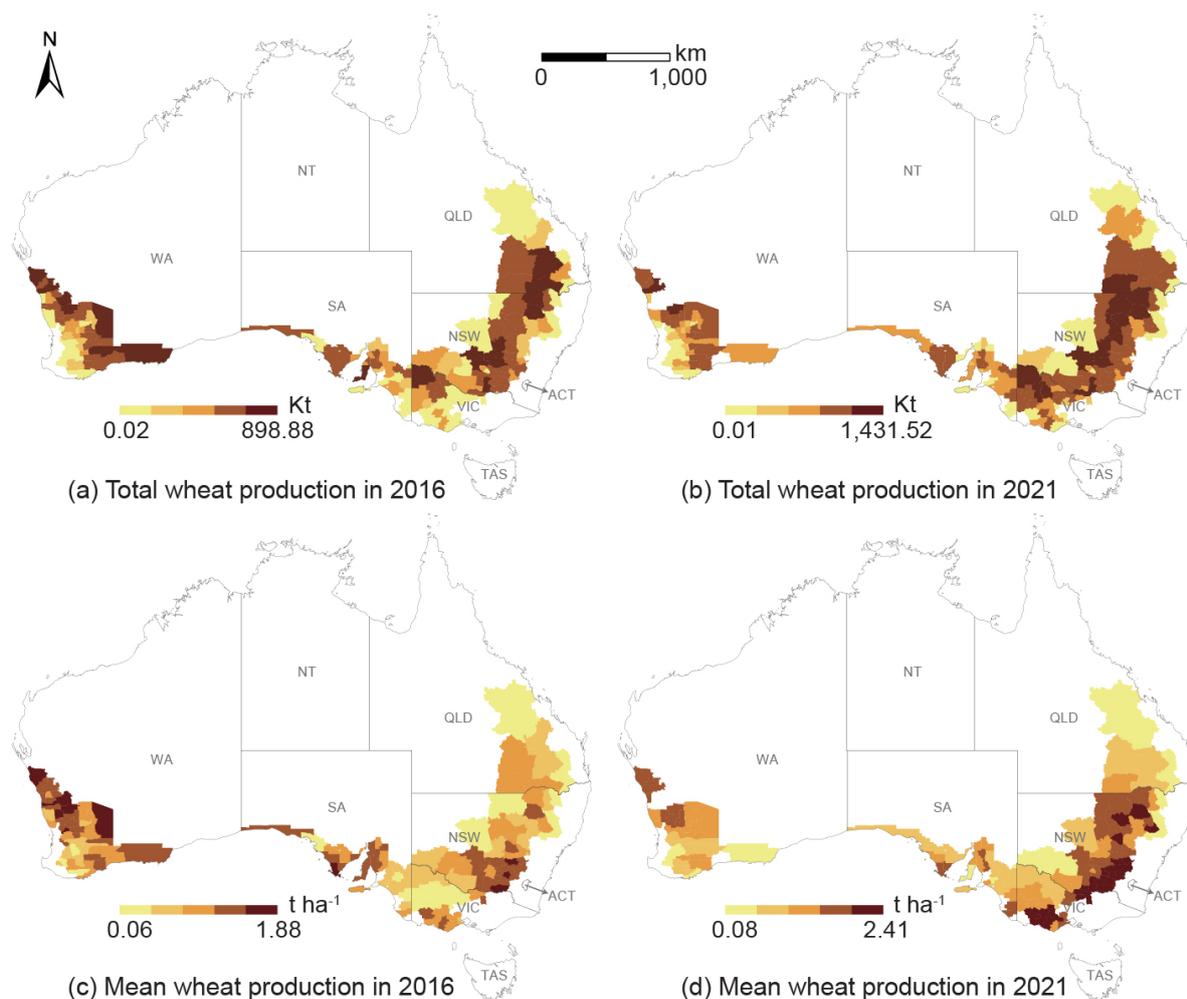


图 3-2 2016 年和 2021 年澳大利亚小麦总产量及平均产量的分布
 Fig. 3-2 Distribution of Australian total wheat production and mean wheat production in 2016 and 2021.

占 44.06%。图 3-2 展示了 2016 与 2021 年澳大利亚小麦带内的总产量与平均产量分布，其空间聚集主要位于中东部、东南沿海及西部至西南部地区，具体高产区包括昆士兰州 (QLD) 东南部、新南威尔士州 (NSW) 中部、维多利亚州 (VIC) 西部、南澳大利亚州 (SA) 东部及西澳大利亚州 (WA) 东南部。

为提高空间表达精度，本章将地方政府区 (Local Government Area, LGA) 与 ABS 定义的小麦带边界叠置，构建修正的空间单元，以避免因过大非生产区导致的偏差 (Feng et al. 2022)。经此方法，2016 年共划定 186 个 LGA，2021 年为 179 个 (Australian Bureau of Statistics 2020, 2021a)。

考虑到各州气候条件与政策差异较大，导致播种、田管及生育期不尽相同，表 3-1 给出了主要州份的小麦生长季。本章按各 LGA 所在州的生长季对应时间，提取对应生长季内的 LGA 区域上的小麦平均产量，以减弱区域面积差异对地理空间分析结果的影

响。

表 3-1 澳大利亚各州的小麦生长季节

Tab. 3-1 The wheat growing seasons in various states of Australia.

| 州 State | 小麦生长季节 Wheat growing seasons |
|-----------------------------|--|
| 新南威尔士州 New South Wales | 4 月至 12 月 From April to December. (Gomez-Macpherson and Richards 1995; Pang et al. 2022; Wang et al. 2015) |
| 维多利亚州 Victoria | 4 月至次年 1 月 From April to January of the following year. (Pang et al. 2022; Wang et al. 2009) |
| 昆士兰州 Queensland | 3 月至 11 月 From March to November. (Cammarano et al. 2012; Obanor et al. 2013) |
| 南澳大利亚州 South Australia | 4 月至 12 月 From April to December. (Pang et al. 2022; Wang et al. 2009) |
| 西澳大利亚州 Western Australia | 5 月至次年 1 月 From May to January of the following year. (Duncan et al. 2017; Shen and Evans 2021) |

3.3.1.2 解释变量数据

本章选取四个大类共 20 个解释变量——地理、气候、土壤属性与环境变量——用以描述澳大利亚小麦产量异质性的主要驱动因素（表 3-2）。变量筛选遵循三项标准：一是与温带作物体系的经验相关性 (Han et al. 2020; Millar et al. 2018)；二是覆盖关键生长过程（温度、降水、土壤水分及质地、植被活力与耗水）；三是 LGA 尺度的数据完整性，统一来源于 Google Earth Engine (MODIS EVI、NDVI) 和各类机构公开的数据集 (ECMWF 再分析、国家土壤调查)。尽管厄尔尼诺-南方涛动和印度洋偶极子等气候模态会影响降水和温度，其局地效应已被本章的降水、温度、土壤湿度与植被指数所捕获，因此未单独纳入。图 3-3 展示了这些变量的空间分布及与产量格局的关系。

3.3.1.3 地理变量

地形直接影响作物所受光照、辐射、温度、降水与气压等多种因素 (Ajami et al. 2020; Kitchen et al. 2003)。本章采用数字高程模型 (DEM) 作为地理变量，数据来自 GEE 中 Geoscience Australia 提供的 DEM (Geoscience Australia 2015)，源于美国宇航局 (NASA) 的 SRTM 数据 (2000 年 2 月获取)，分辨率 30 m，并经平滑处理以降低噪声、提升地表形态表达。接着，基于 DEM 进一步计算了坡度与坡向。

表 3-2 潜在影响小麦生产空间差异的解释变量概述

Tab. 3-2 A summary of explanatory variables that potentially affect spatial disparities of wheat production.

| 类别 Category | 变量 Variable | 代码 Code | 数据产品 Product | 分辨率 Resolution |
|-------------------|---|------------|---------------------|-------------------|
| 地理 Geography | 海拔 Elevation | EL | DEM-S | 30 m |
| | 坡度 Slope | SLP | DEM-S | 30 m |
| | 坡向 Aspect | ASP | DEM-S | 30 m |
| 气候 Climate | 气温 Air temperature | AT | ERA5_land | 0.25° |
| | 总降水量 Total precipitation | TP | ERA5_land | 0.25° |
| | 地表气压 Surface pressure | SP | ERA5_land | 0.25° |
| | 风速 Wind speed | WS | ERA5_land | 0.25° |
| 土壤数据 Soil data | 有效持水量 Available water capacity | AWC | CSIRO/SLGA | 92.77 m |
| | 容重 (全土层) Bulk density (Whole earth) | BDW | CSIRO/SLGA | 92.77 m |
| | 黏土含量 Clay | CLY | CSIRO/SLGA | 92.77 m |
| | 总氮 Total Nitrogen | NTO | CSIRO/SLGA | 92.77 m |
| | 总磷 Total Phosphorus | PTO | CSIRO/SLGA | 92.77 m |
| | 粉粒 Silt | SLT | CSIRO/SLGA | 92.77 m |
| | 砂粒 Sand | SND | CSIRO/SLGA | 92.77 m |
| | 土壤有机碳 Soil organic carbon | SOC | CSIRO/SLGA | 92.77 m |
| | pH ($CaCl_2$) | pHc | CSIRO/SLGA | 92.77 m |
| | pH ($CaCl_2$) | | | |
| 环境 Environment | 蒸散量 Evapotranspiration | ETa | CMRSET Landsat V2.2 | 30 m |
| | 净初级生产力 Net primary production | NPP | MOD17A3HGF V6.1 | 500 m |
| | 归一化植被指数 Normalized difference vegetation index | NDVI | MOD13A2 V6.1 | 1000 m |
| | 增强型植被指数 Enhanced vegetation index | EVI | MOD13A2 V6.1 | 1000 m |

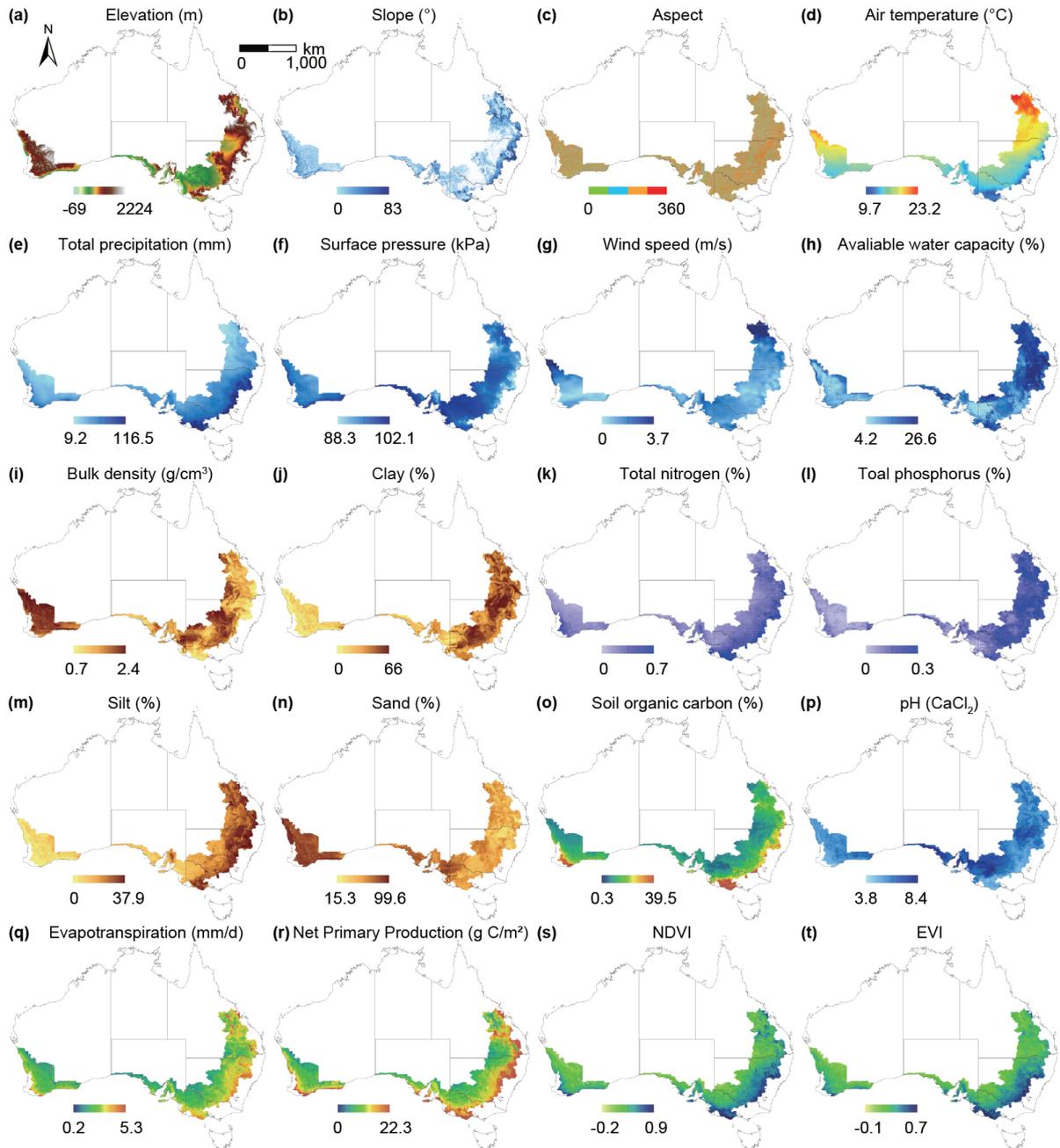


图 3-3 影响小麦产量的解释变量的空间分布 (2021 年)。(a) - (c) 地理变量, (d) - (g) 气候变量, (h) - (p) 土壤属性, (q) - (t) 环境变量

Fig. 3-3 Spatial distributions of explanatory variables in 2021. (a) - (c) Geographical variable, (d) - (g) Climate variables, (h) - (p) Soil attributes, (q) - (t) Environmental variables

3.3.1.4 气候变量

选取气温 (AT)、总降水量 (TP)、地表气压 (SP) 和风速 (WS) 四项气候变量, 均来源于 ECMWF 第五代全球气候再分析产品 ERA5(Hersbach et al. 2023), 空间分辨率 0.25° 。这些变量直接影响作物生长的热量与水分条件, 是产量空间差异的主要驱动因素(Song et al. 2019; Zhang et al. 2022)。

3.3.1.5 土壤属性

土壤属性通过影响养分供给、持水能力、质地和 pH 等，对作物生长发育产生显著作用 (Kitchen et al. 2003; Miller et al. 1988; Nabiollahi et al. 2020)。本章使用 GEE 中的澳大利亚土壤与景观网格 (SLGA) (Rossel et al. 2015)，分辨率 3" (约 90 m)。选取 9 个土壤属性：有效持水量 (AWC)、容重 (BDW)、黏土 (CLY)、全氮 (NTO)、全磷 (PTO)、粉粒 (SLT)、砂粒 (SND)、土壤有机碳 (SOC) 及 $\text{pH}(\text{CaCl}_2)$ (pHc)。各变量提供 0–15、5–15、15–30、30–60、60–100 与 100–200 cm 六层深度信息。

3.3.1.6 环境变量

环境状况通过蒸散量 (ETa)、净初级生产力 (NPP)、归一化植被指数 (NDVI) 与增强型植被指数 (EVI) 表征 (Qader et al. 2018; Wang et al. 2019; Xu et al. 2019)。ETa 来源于 CSIRO Land and Water (CMRSET Landsat V2.2, 分辨率 30 m) (Guerschman et al. 2022); NPP、NDVI 与 EVI 均来自 MODIS 产品，通过 GEE 获取 (Didan 2021; Running and Zhao 2021)。

其中，NPP 与 NDVI 等植被指数在一定程度上能够直接反映作物生长状况与产量变化趋势，因此可视为产量的表征因子。然而，在本研究的空间预测与异质性分析中，将其与气候、土壤等影响因子一并纳入模型，旨在利用多源变量的互补性，不仅保留植被指数对产量的直接表征作用，同时结合土壤肥力、气候条件等因素，更全面、系统地解释产量空间差异性。这些环境与植被指标既反映了区域生态健康水平，也为揭示产量分布的多维驱动机制提供了重要支撑。

3.3.1.7 数据预处理

在模型构建前，需对收集的原始数据进行系统性预处理，以满足模型计算的输入要求。预处理主要包括以下四个方面：首先，统一所有解释变量的空间单位，确保空间尺度一致性；其次，对数据进行标准化处理，计算各地方政府区域 (LGA) 的均值，并将温度数据由开尔文 (Kelvin) 转换为摄氏度 (Celsius)；第三，进行异常值识别与剔除，尤其在土壤容重 (BDW) 与土壤有机碳 (SOC) 两个变量中，存在个别像元值远高于周边值的情况。本研究设定阈值为均值的 2.5 倍标准差，对超出此范围的数值予以剔除；最后，鉴于土壤属性数据包含六个不同土层的指标，采用主成分分析 (Principal Component Analysis, PCA) 方法对其进行降维处理，提取主要成分，以保留最具信息量的特征并提升后续建模效率。表 3-3 总结了各类解释变量的预处理过程。此外，GOZH 模型在计算过程中自动检测输入变量间的多重共线性，因此在数据预处理阶段无需进行多重共线性校正。

表 3-3 解释变量数据的预处理方法与流程

Tab. 3-3 Methods and procedures for pre-processing explanatory variable data.

| 预处理方法 Pre-processing | 变量 Variable |
|--|---|
| 计算每个 LGA 的平均值 Average to each LGA. | DEM, SLP, ASP |
| 计算每个 LGA 的年度平均值 Calculate the annual average for each LGA. | NPP |
| 基于不同州的小麦生长季计算 LGA 内的平均值，并转换单位为摄氏度 (°C) Calculate the average value within the corresponding LGA based on the wheat growing season in different states, and convert the unit to degree Celsius (°C). | AT |
| 进行离群点分析和主成分分析，并计算每个 LGA 的平均值 Outlier analysis and principal component analysis are performed, and the average is calculated for each LGA. | BDW, SOC |
| 进行主成分分析，并计算每个 LGA 的平均值 Conduct the principal component analysis, and average to each LGA. | AWC, CLY, NTO, PTO, SLT, SND, pHc |
| 基于不同州的小麦生长季计算 LGA 内的平均值 Calculate the average value within the corresponding LGA based on the wheat growing season in different states. | TP, WS, ETa, NDVI, EVI |

3.3.2 莫兰指数和局部空间自相关分析

为识别小麦产量的空间分布模式与聚集性特征，本研究采用了空间自相关分析方法，包括全局 Moran's I 指数和局部空间自相关分析 (Local Indicators of Spatial Association, LISA)。

莫兰指数 (Moran's I) 是地统计学中广泛应用的一种空间自相关性度量指标，用于衡量空间数据的自相关性，其值范围为 $[-1, 1]$ 。当 Moran's I 值接近 1 时，表示空间数据存在显著的正相关性；当接近 -1 时，表示存在显著的负相关性；接近 0 则表示空间数据无明显自相关性；等于 0 则表示空间分布为随机状态，无显著自相关性 (Li et al. 2007; Moran 1950)。其计算公式如下：

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (3-1)$$

式中， n 为样本数量， x_i 与 x_j 分别为第 i 与 j 个空间单元的属性值， \bar{x} 为变量的平均值， w_{ij} 为第 i 与 j 个单元之间的空间权重。

局部空间自相关分析 (Local indicators of spatial association, LISA) 则用于识别空间数据中的热点和冷点区域，以及空间数据的聚集性和显著性水平。LISA 主要用于局

部尺度上分析空间数据的聚集性与相关性，是 Moran's I 的局部版本 (Anselin 1995)。该方法通过比较单个空间单元与其邻近单元之间属性值的协方差与方差比，判断其空间聚集或离散情况 (Anselin 2005)。其计算公式如下：

$$I_i = \frac{x_i - \bar{x}}{s^2} \sum_{j=1}^n w_{ij}(x_j - \bar{x}) \quad (3-2)$$

式中， I_i 为局部 Moran's I 值， x_i 与 x_j 分别为第 i 与 j 个单元的属性值， \bar{x} 为均值， s^2 为变量 x 的方差， w_{ij} 为空间权重矩阵。

本研究中采用的空间权重计算方法为高斯核函数 (Gaussian Kernel function)，并将对角线元素权重设定为 1。

3.3.3 传统的地理探测器模型

地理探测器(王劲峰, 徐成东 2017) 通过“因子探测器”“交互探测器”“风险探测器”与“生态探测器”四类统计模块完整刻画因子—响应体系 (表 3-4)。设研究区包含 N 个空间单元，响应变量记为 $Y = \{y_i\}_{i=1}^N$ ，其中 y_i 为第 i 个空间单元的观测值；解释变量 X 经离散化后形成 L 个空间分层，第 h 层包含 N_h 个样本。其核心统计量 q 值定义为

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2}, \quad (3-3)$$

式中 σ_h^2 为第 h 层响应变量的方差， σ^2 为全域响应变量的方差。 N 为总样本量， N_h 为第 h 层样本量。当 $q \rightarrow 1$ 时表明 X 几乎完全决定 Y 的空间格局； $q \rightarrow 0$ 则说明二者空间分布独立。式 (3-3) 亦可转化为非中心 F 分布，以实施显著性检验：

$$F = \frac{N - L}{L - 1} \frac{q}{1 - q} \sim F(L - 1, N - L; \lambda), \quad (3-4)$$

其中 $F(\cdot)$ 表示 F 分布， λ 为非中心参数。

若需比较两解释变量 X_1 与 X_2 的协同作用，可分别计算 $q(X_1)$ 、 $q(X_2)$ 以及叠加分层后的 $q(X_1 \cap X_2)$ ；当 $q(X_1 \cap X_2)$ 大于 $q(X_1) + q(X_2)$ 时，说明二者存在非线性增强效应。风险探测器通过

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\text{Var}(\bar{Y}_1)/n_1 + \text{Var}(\bar{Y}_2)/n_2}} \quad (3-5)$$

比较不同子区均值，以识别高风险与低风险区域；式中 \bar{Y}_1 和 \bar{Y}_2 分别为两类分区的均值， n_1 、 n_2 分别为样本量， $\text{Var}(\bar{Y}_1)$ 和 $\text{Var}(\bar{Y}_2)$ 为均值的方差估计。生态探测器则利用

$$F = \frac{N_{X_1}(N_{X_2} - 1) \text{SSW}_{X_1}}{N_{X_2}(N_{X_1} - 1) \text{SSW}_{X_2}} \quad (3-6)$$

来检验不同解释变量对 Y 的作用强度是否存在显著差异，其中 N_{X_1} 和 N_{X_2} 分别为变

量 X_1 与 X_2 的分层数, SSW 为层内平方和。

表 3-4 GeoDetector 四类探测器功能摘要

| 探测器 | 主要统计量 | 分析目标与解释 |
|-------|-------------------|------------------------|
| 因子探测器 | q (式 (3-3)) | 评估单因子对响应变量空间方差的解释比例 |
| 交互探测器 | $q(X_1 \cap X_2)$ | 判断两因子叠加后对 Y 的协同或抑制效应 |
| 风险探测器 | t 统计量 | 比较不同分层均值, 识别高/低风险区 |
| 生态探测器 | F 统计量 | 检验不同因子对 Y 影响强度是否显著差异 |

在实际应用中, 连续型解释变量通常先以等距、分位数或自然断点等方法离散化为 3-7 个等级; 随后通过遍历不同分层方案, 选取 q 值最大的方案以保证分区最优。GeoDetector 支持栅格、面和点等多种数据类型, 并已实现于 R 语言等多种软件平台; 在典型环境与健康风险研究中, 该方法往往能以较少样本实现对复杂空间过程的稳健解释, 展示了其在高维共线样本、非线性交互识别及局地热点检测方面的综合优势。

3.3.4 地理最优区域异质性 (GOZH) 模型

上节已经详细介绍了地理探测器的原理。其中, 因子探测 (Factor Detection) 是核心模块, 可通过计算解释力 (Power of Determinants, PD) 值, 衡量各解释变量对响应变量空间分布的解释能力。其计算公式如下 (Wang et al. 2010):

$$PD = 1 - \frac{\sum_{z=1}^h N_z \sigma_z^2}{N \sigma^2} = 1 - \frac{SSW}{SST} \quad (3-7)$$

式中, SSW 和 SST 分别表示区域分层内的平方和与由解释变量确定的总体平方和; N_z 和 σ_z 表示地理区域 z ($z = 1, \dots, h$) 内的样本数量与标准差; N 与 σ 分别为整个研究区域的样本数量与标准差。PD 值的取值范围为 $[0, 1]$, 该值用于衡量解释变量对响应变量的解释力。PD 值越高, 说明解释变量与响应变量之间的关联性越强, 对空间分布差异的解释能力越显著。

GOZH (Geographically Optimal Zones-based Heterogeneity) 模型旨在识别地理空间上最优的区域划分, 使区域间差异最大化, 区域内相似性最小化, 从而更有效地揭示影响因子的空间异质性特征 (Luo et al. 2022)。该模型通过基于决策树的空间分层优化方法, 有效地估计多个解释变量之间的交互关系, 从而识别影响空间分布的决定因子。

此外, GOZH 模型进一步提升了多变量交互作用估计的稳定性与可靠性。在本研究中, GOZH 模型不仅定义了传统的 PD 值用于解释变量与空间分区之间的关联性, 还提出了最优解释力 (Optimal Power of Determinants, OPD) 值, 用于表示解释变量在最优地理区域划分下对响应变量的最大解释能力 (Luo et al. 2022):

$$\Omega = \max(PD) = \gamma(X, D) = 1 - \frac{\min(SSW_{X,D})}{SST} \quad (3-8)$$

式中, X 表示一个或多个解释变量, D 为描述地理区域的分层变量, $SSW_{X,D}$ 为由解释变量 X 所定义的地理分层区域内的平方和。 Ω 值表示解释变量在最优区域划分下的最大 PD 值, 即 OPD 值。

GOZH 模型可量化各解释变量对小麦产量空间异质性的贡献。为防止过度分割并确保地理区域划分的合理性, 模型构建了基于最小分裂样本量限制的决策树 (minsplit = 10), 即每个节点至少包含 10 个 LGA 样本 (Breiman 1984)。在逐步空间离散化的过程中, 对包含 LGA 数量不足 10 个的类别不再继续划分。在每一轮划分中, 计算每个解释变量对应类别的 PD 值, 即 Ω 值, 以揭示其对小麦产量空间分布的影响和贡献程度。

最优地理分区应具备“区域内方差最小、区域间方差最大”的特性。为计算 Ω , 需最小化 $SSW_{X,D}$, 其具体计算公式如下:

$$\min(SSW_{X,D}) = \min \left\{ \sum_{z=1}^h \sum_{j=1}^{N_z} (y_{zj} - \bar{c}_z)^2 \right\} \quad (3-9)$$

式中, y_{zj} 表示区域 z 中第 j 个观测值 ($j = 1, \dots, N_z$), \bar{c}_z 为区域 z 中小麦产量的均值。为求解上述公式, 需要通过解释变量对小麦产量数据进行逐步空间离散化, 选择最优的划分变量及其断点。 Ω 的详细推导过程参见 Luo 等人发表的研究成果 (Luo et al. 2022)。

该空间离散化过程与经典分类与回归树 (Classification and Regression Tree, CART) 算法具有高度相似性 (Breiman 1984)。GOZH 模型主要依托 R 语言中的“rpart”包 (Therneau and Atkinson 2023) 和“GD”包 (Song et al. 2020) 来实现。

3.3.4.1 基于最优参数的地理探测器 (OPGD) 模型

上述已经提到, 地理探测器是一种用于揭示空间异质性及其驱动机制的统计方法 (Wang and Xu 2017), 而基于最优参数的地理探测器 (Optimal Parameters-based Geographical Detector, OPGD) 模型是在传统地理探测器模型基础上的一种优化扩展方法。该模型通过系统化地优化变量离散化方式、分层数及空间尺度参数, 提升了模型对空间异质性的解释力和稳健性 (Song et al. 2020)。

与传统的地理探测器模型类似, OPGD 模型也主要包括五个模块: 因子探测、参数优化、交互作用探测、风险探测和生态探测。其中核心的因子探测模块用于评估某解释变量对因变量空间分布的解释力, 其计算公式如下:

$$Q_v = 1 - \frac{\sum_{j=1}^L N_{v,j} \cdot \sigma_{v,j}^2}{N_v \cdot \sigma_v^2} \quad (3-10)$$

式中， Q_v 表示解释变量 v 对因变量空间异质性的解释力， L 为该变量被划分的层数， N_v 和 σ_v^2 分别为整个研究区域的样本数量与方差， $N_{v,j}$ 和 $\sigma_{v,j}^2$ 分别表示第 j 个子区域中的样本数量和因变量的方差。 Q_v 的取值范围为 $[0, 1]$ ，值越大表示变量解释能力越强。

与传统地理探测器依赖经验划分不同，OPGD 对每个连续变量在多种离散化方法（如等距、分位数、自然断点等）、不同分类数（如 3–7 类）及空间尺度参数组合下计算其 Q 值，最终选取解释力最强的组合作为最优参数，以避免人为主观性带来的偏差。

在此基础上，OPGD 还引入交互作用探测模块，通过比较两个变量交互后的 $Q_{u \cap v}$ 与其单独 Q_u 、 Q_v 之间的关系，用以判断变量间交互的增强、减弱或独立关系。进一步的风险探测与生态探测模块则用于识别不同分层之间因变量的平均差异及其统计显著性。

OPGD 模型可适用于点状、面状与栅格等多种空间数据类型，并已集成于 R 语言的 GD 软件包中，支持快速计算和可视化输出，适合大样本数据的高效分析。

本研究采用 OPGD 模型对解释变量的空间分层异质性进行度量，并与本文提出的 GOZH 模型进行对比，以验证后者在复杂空间结构识别中的优势。

3.3.5 识别变量之间的交互影响

GOZH 模型是通过逐步空间离散化以及多变量之间的交互作用，实现对地理最优区域的识别。GOZH 模型以四类解释变量为输入，系统性地探讨这些因素对小麦产量空间差异性的影响机制及其相互关系。通过比较各类因子在划分地理最优区域中的贡献程度，明确识别出主导影响空间异质性的变量类别，从而揭示空间差异性背后的驱动机制。

在小麦产量空间差异性案例研究中，采用 GOZH 模型对具有小麦产量的 LGA 区域进行空间离散化，并利用 20 个解释变量构建二叉树结构。在迭代离散化过程中，LGA 被划分为多个地理最优区域。在这些区域内部，属于同一类别的区域表现出显著的同质性，而不同类别之间则呈现出显著的异质性。

随后，GOZH 模型被应用于多个空间尺度上，使用统一的输入与输出变量，以保证结果的可比性。应用范围涵盖整个澳大利亚大陆、热点区与非热点区，以及各个州级行政单元。通过对比不同区域所获得的 Ω 值，评估模型在不同地理尺度下的稳健性与适应性，从而深入理解 GOZH 模型的尺度依赖性表现及其在多空间层级下的适用性。

3.3.6 模型验证与评估方法

为验证所构建的地理空间机器学习模型的有效性，本文引入另一种基于参数优化的地理探测器模型 (OPGD) 进行对比分析。OPGD 模型是在经典地理探测器模型 (Geographical detector, GD) 基础上发展而来 (Song et al. 2020)，其核心优势在于优化了空间数据的离散化与尺度选择过程，从而确定最优的参数组合以增强模型性能。该优化过程有助于提取空间解释变量中所蕴含的关键地理特征，提升对空间异质性的识别能力。

通过与 OPGD 模型结果的对比，可以验证和评估 GOZH 模型在分析澳大利亚小麦产量空间异质性方面的有效性和可靠性。OPGD 模型在 R 语言环境下通过“GD”软件包实现。通过比较两个模型下各解释变量的 PD 值及其变化幅度，进一步评估 GOZH 模型在空间异质性分析中的优势表现。此外，还分析了四个主导变量在不同空间分层中的 PD 值变化趋势，以此揭示 GOZH 模型与 OPGD 模型在空间尺度识别能力方面的差异。

3.4 结果分析

3.4.1 澳大利亚小麦产量的空间分布模式分析

图 3-4 展示了澳大利亚两个年份小麦产量的空间分布格局。从图 3-4 (a) 和 (d) 可见，2016 年和 2021 年小麦产量的 Moran's I 值分别为 0.532 和 0.669，表明小麦产量在空间上存在显著聚集性。图 3-4 (b) 和 (e) 标识出高产区（热点）与低产区（冷点），进一步揭示了小麦产量的空间聚集特征。2016 年共有 41 个热点地区，主要分布在新南威尔士州 (NSW) 中南部、南澳大利亚州 (SA) 南部以及西澳大利亚州 (WA) 中西部和南部，占研究区域地方政府区域 (LGA) 的 22%。到 2021 年，热点数量增加至 44 个，主要集中在 NSW 和维多利亚州 (VIC)，占研究区域的 24.6%。热点分布区域的变化可能与气候和环境条件的变动密切相关，例如降水变化、气温波动以及厄尔尼诺-南方涛动 (ENSO) 现象的影响。

图 3-4 (c) 和 (f) 显示了不同区域在统计学上的显著性水平（如 $p < 0.001$ 、 $p < 0.01$ 、 $p < 0.05$ ），表明热点和冷点区域均具有统计上的显著性。这一结果说明小麦产量在热点区域的空间聚集并非随机现象，其空间自相关性具有显著意义。同样，2016 年与 2021 年的冷点区域也表现出高度显著性，反映出低产区可能受持续性不利因素的影响，如土壤质量较差或气候条件不利。

图 3-5 展示了热点、冷点及其他区域的小麦产量统计特征。箱线图表明，热点区域的小麦产量中位数较高，分别为 2016 年的 $1.05 t \cdot ha^{-1}$ 和 2021 年的 $1.72 t \cdot ha^{-1}$ ，反映出热点区域的产量水平始终优于其他地区。2021 年中位产量较 2016 年提升了 63.8%。

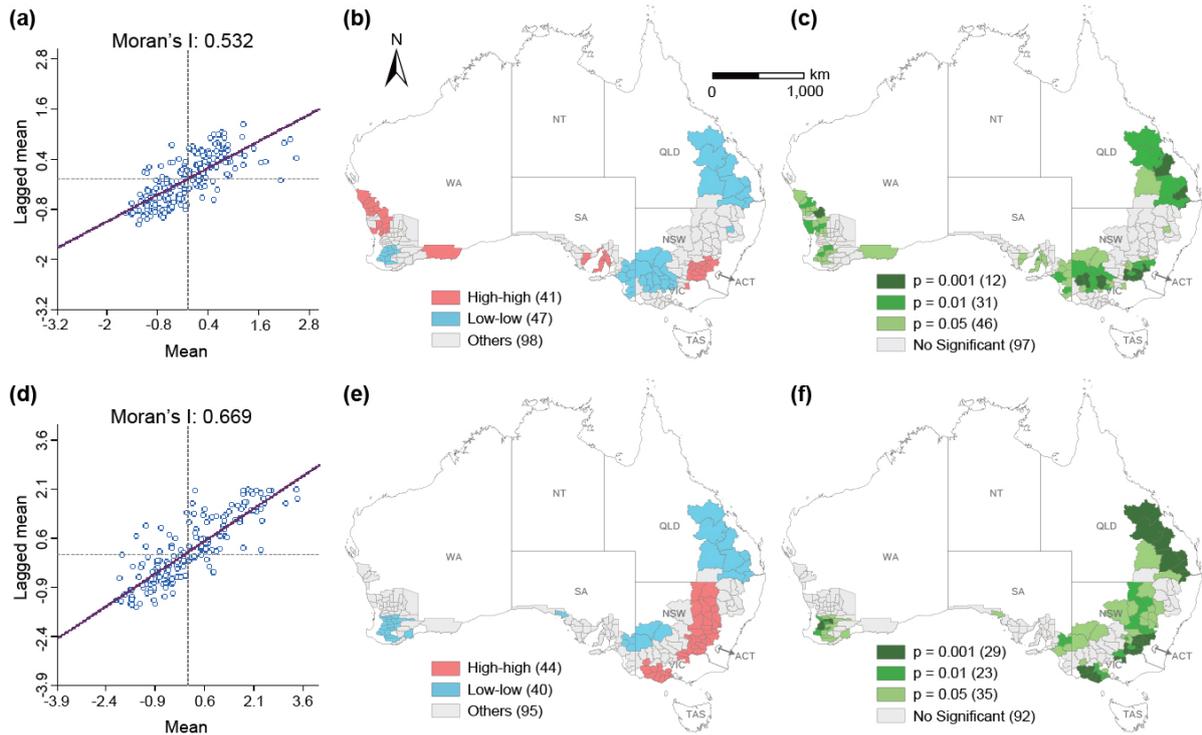


图 3-4 Moran's I 与 LISA 分析结果。2016 年：(a) Moran's I 值，(b) 基于 LISA 的热点分析，(c) 基于 LISA 的显著性分析；2021 年：(d) Moran's I 值，(e) 基于 LISA 的热点分析，(f) 基于 LISA 的显著性分析。

Fig. 3-4 Results of Moran's I and LISA analysis. 2016: (a) Value of Moran's I, (b) hotspot analysis based on LISA, (c) significance analysis based on LISA; 2021: (d) Value of Moran's I, (e) hotspot analysis based on LISA, (f) significance analysis based on LISA.

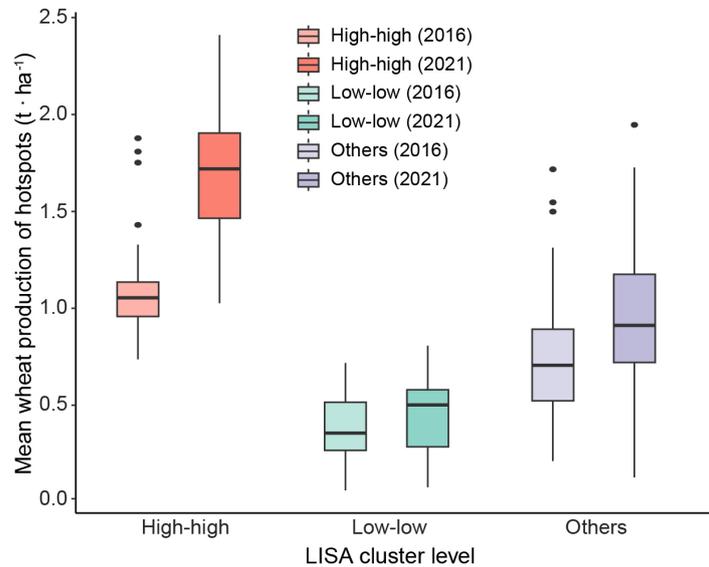


图 3-5 基于 LISA 分析结果的热点区域与非热点区域的小麦平均产量

Fig. 3-5 Mean wheat production of hotspot and non-hotspot areas based on the results of LISA analysis.

2016 年热点区域的四分位距较小，产量变异性较低；而 2021 年则展现出更大的产量波动。此外，2016 年部分 LGA 表现出异常高产，可能受局部优越条件影响。相比之下，冷点区域的中位产量显著偏低，而其他区域则处于两者之间，四分位距适中，产量相对稳定，波动性较小。

表 3-5 澳大利亚不同州小麦产量的局部空间自相关 (LISA) 聚类统计分析

Tab. 3-5 Local Indicators of Spatial Association (LISA) clustering statistics for wheat production across different states in Australia.

| | | LISA Statistics | NSW | VIC | WA | SA | QLD | Australia |
|------|-----------|-----------------|-------|-------|-------|-------|------|-----------|
| 2016 | High-high | Count | 13 | 1 | 18 | 9 | - | 41 |
| | | P_{sum}^1 | 15.37 | 1.07 | 19.43 | 9.15 | - | 45.03 |
| | | P_{mean}^2 | 1.18 | 1.07 | 1.08 | 1.02 | - | 1.08 |
| | Low-low | Count | 4 | 12 | 12 | 8 | 11 | 47 |
| | | P_{sum} | 2.29 | 3.56 | 5.01 | 3.58 | 3.56 | 18.01 |
| | | P_{mean} | 0.57 | 0.30 | 0.42 | 0.45 | 0.32 | 0.38 |
| | Others | Count | 30 | 13 | 35 | 19 | 1 | 98 |
| | | P_{sum} | 19.47 | 9.52 | 27.31 | 15.12 | 0.69 | 72.11 |
| | | P_{mean} | 0.65 | 0.73 | 0.78 | 0.80 | 0.69 | 0.74 |
| 2021 | High-high | Count | 30 | 14 | - | - | - | 44 |
| | | P_{sum} | 48.68 | 25.01 | - | - | - | 73.69 |
| | | P_{mean} | 1.62 | 1.79 | - | - | - | 1.67 |
| | Low-low | Count | 2 | 1 | 20 | 6 | 11 | 40 |
| | | P_{sum} | 0.85 | 0.68 | 9.95 | 3.11 | 3.94 | 18.53 |
| | | P_{mean} | 0.42 | 0.68 | 0.50 | 0.52 | 0.36 | 0.46 |
| | Others | Count | 15 | 11 | 38 | 30 | 1 | 95 |
| | | P_{sum} | 14.34 | 9.84 | 36.35 | 26.36 | 1.11 | 88.00 |
| | | P_{mean} | 0.96 | 0.89 | 0.96 | 0.88 | 1.11 | 0.93 |

¹ Sum of the mean production ($t \text{ ha}^{-1}$).

² Average of the mean production ($t \text{ ha}^{-1}$).

表 3-5 汇总了不同区域（热点、冷点和其他区域）LGA 数量、总平均产量及其产量特征。2021 年新南威尔士州 (NSW) 的热点区域数量最多，达 30 个，占有热点区域的 68.2%。这些区域的平均产量为 $1.62 t \cdot \text{ha}^{-1}$ ，明显高于 2016 年 NSW 热点的 $1.18 t \cdot \text{ha}^{-1}$ 。此外，WA 冷点所占比例由 2016 年的 25.5% 上升至 2021 年的 50%，可能与该地区面临的干旱条件有关。维多利亚州 (VIC) 的热点数量则从 2016 年的 1 个增加至 2021 年的 14 个，比例由 2.4% 上升至 31.8%；同时冷点数量由 12 个锐减至 1 个，表明该州小麦产量呈现显著增长趋势。

综上所述，澳大利亚小麦主产区呈现出明显的空间分布特征和聚集模式，表明研

究区域具有良好的农业生产条件。小麦产量的空间分布格局有助于我们深入理解其空间聚集性和自相关特征，以及不同区域之间的产量差异。这些发现为进一步分析小麦产量的空间差异性及其影响因素提供了科学依据。

3.4.2 单个变量对小麦产量分布的贡献度

GOZH 通过对澳大利亚小麦主产区的平均产量数据及其解释变量进行模拟，可以识别小麦产量的空间异质性格局及其决定性因素。本部分内容计算了单个变量对小麦产量空间差异性的贡献度。

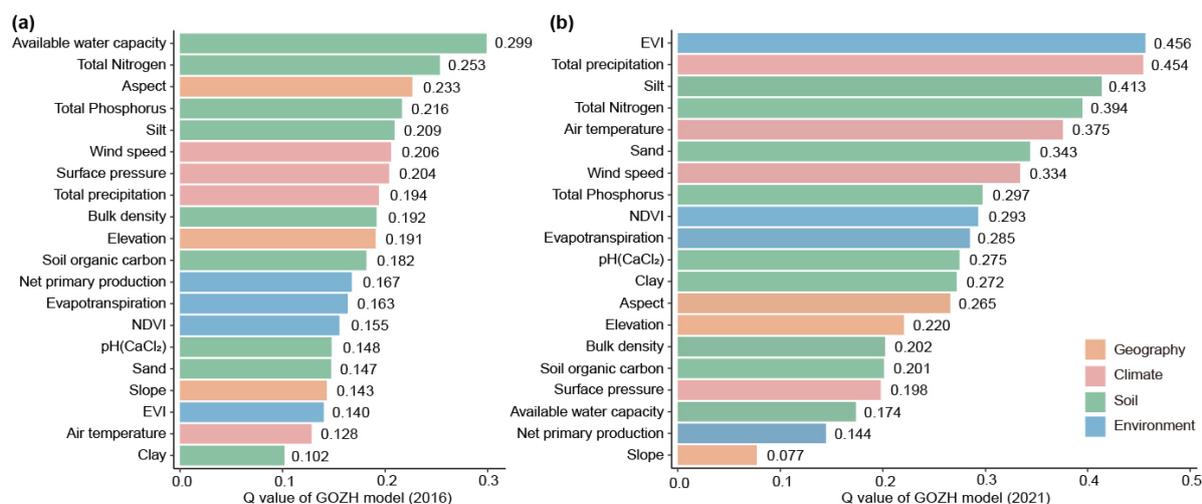


图 3-6 基于地理最优分区异质性 (GOZH) 模型评估小麦产量解释力 (PD) 的结果

Fig. 3-6 Results of the geographical optimal zones-based heterogeneity (GOZH) model for assessing power of determinants (PD) of wheat production.

图 3-6 展示了通过 GOZH 模型评估 2016 年与 2021 年各变量对小麦产量异质性贡献的结果。尽管不同变量对空间格局的影响程度有所差异，但总体上均对小麦产量具有显著影响。解释变量主要分为四大类：地理因素、气候变量、土壤属性和环境条件。

2016 年，土壤属性的影响最为显著，共有九个土壤变量对产量分布产生了重要作用，其中有效持水量 (AWC，贡献率 29.9%) 和全氮 (NTO，贡献率 25.3%) 贡献最高，这与小麦生长对水分保持和养分吸收的依赖密切相关。地形因素中的坡向 (ASP) 也具有较高贡献 (23.3%)，其对光照、辐射以及地表径流的影响直接作用于作物生长。气候变量作为基础性影响因子，风速 (WS，20.6%)、地面气压 (SP，20.4%)、降水量 (TP，19.4%) 和气温 (AT，12.8%) 共同影响作物授粉、光合作用效率、蒸腾速率及水分吸收。此外，环境类变量如净初级生产力 (NPP，16.7%)、蒸散量 (ETa，16.3%)、NDVI (15.5%) 和 EVI (14%) 对产量也具有较强影响，反映出良好的植被生长环境有助于提高光合作用效率及产量。

2021 年，各变量的影响出现明显变化。尽管主要影响因子未发生本质性变化，但

EVI (45.6%)、TP (45.4%) 与 SLT (41.3%) 成为前三大决定性变量，均贡献超过 40%，表明该三类变量在 2021 年对小麦产量空间格局具有关键影响。土壤属性持续发挥重要作用，反映出澳大利亚耕地多分布于沿海及丘陵地区，土壤结构、水分保持能力、养分含量、有机质含量及 pH 值对作物生长均有深远影响。气候变量方面，TP (45.4%)、AT (37.5%)、WS (33.4%) 和 SP (19.8%) 在 2021 年均具有显著贡献，尤其 TP 和 AT 的影响进一步印证了小麦对降水与气温高度敏感的特点。EVI 的显著影响也再次强调了植被覆盖对小麦生长的重要性。

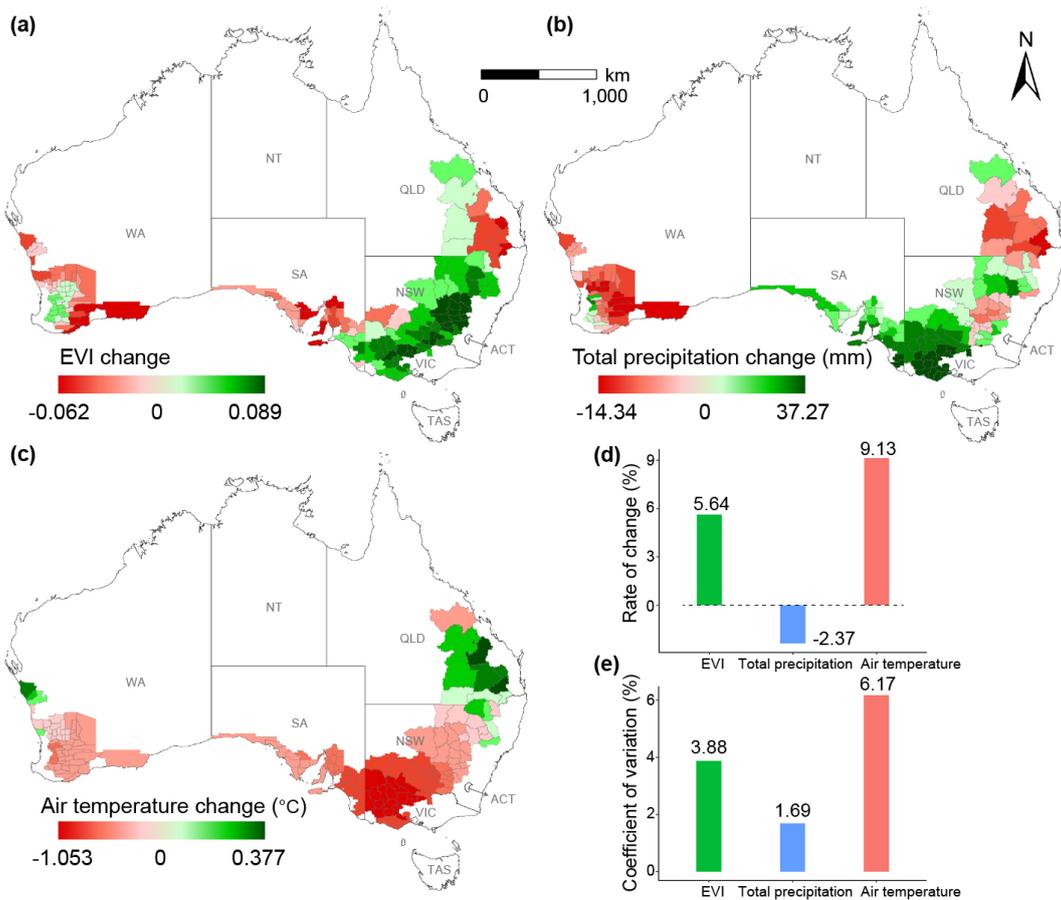


图 3-7 2016 至 2021 年影响小麦产量空间差异的前三个决定因子的变化及其变化率与变异系数
Fig. 3-7 Changes in the top three determinants affecting spatial disparities in wheat production from 2016 to 2021, as well as their change rates and coefficients of variation.

图 3-7 描述了 2016 至 2021 年间主要决定性因子的变化趋势。图 3-7 (a) 显示 EVI 在昆士兰州 (QLD)、NSW 和 VIC 显著增加，WA 西南部略有上升，而其他区域呈不同程度下降。图 3-7 (b) 表明 TP 在 NSW、VIC 和 SA 显著增加，最大增幅达 37.27 mm，而 WA 与 QLD 出现明显减少。图 3-7 (c) 显示 QLD 及 WA 西部的气温上升显著，最大升幅为 0.377 °C，其余区域则有小幅下降 (<1 °C)。图 3-7 (d) 和 (e) 分别展示这三类变量的变化率与变异系数，其中气温的变化率 (9.13%) 与变异系数 (6.17%) 最高，说

明 2021 年气温较 2016 年有显著提升；EVI 略有上升，表明全国植被覆盖状况整体改善；TP 则下降了 2.37%，其变异系数为 1.69%，这表明即便是微小的变化亦可能对小麦生长产生重要影响。

3.4.3 识别小麦产量的地理最优区域

GOZH 模型结合决策树与分层分析方法，将研究区域划分为四个地理最优区域。尽管 2016 年与 2021 年所采用的划分变量有所不同，但各年份划分得到的区域在空间上仅表现出轻微差异。该方法通过最大化区域间差异并最小化区域内产量差异，有效揭示了小麦产量的空间异质性特征。

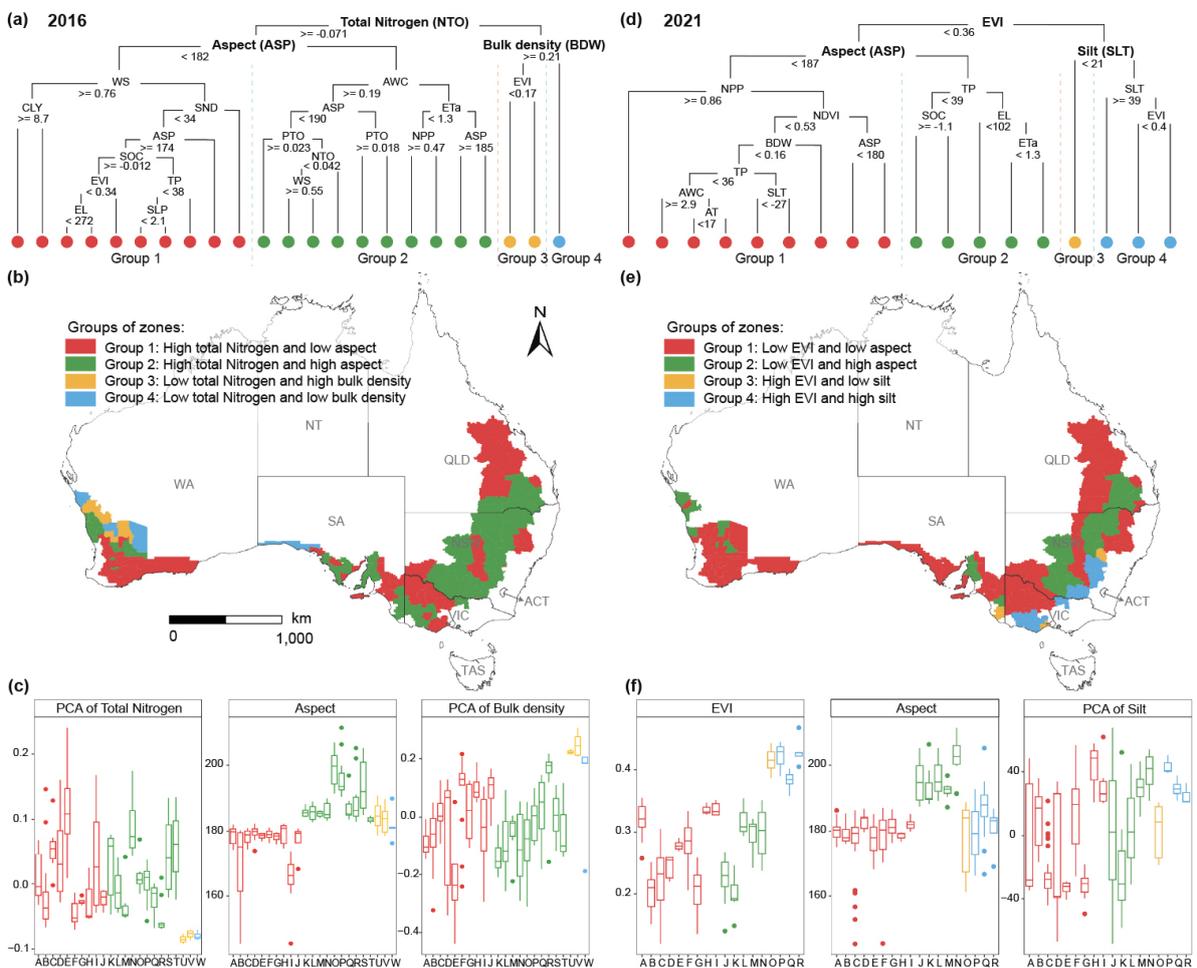


图 3-8 最优区域识别决策树 (a, d)、基于 GOZH 模型识别的 LGA 小麦产量地理最优区域 (b, e)，以及用于解释区域特征的统计摘要 (c, f)

Fig. 3-8 Decision tree of identifying optimal zones (a, d), geographically optimal zones of wheat production at the LGAs identified using the GOZH model (b, e), and statistical summaries of explanatory variables within zones for explaining characteristics of zones (c, f).

图 3-8 展示了利用决策树和分层分析方法对 2016 年和 2021 年小麦主产区进行四大区域划分的过程。2016 年中，首要划分变量为总氮含量 (NTO)，该变量在小麦生长

过程中对养分吸收起关键作用，进而直接影响产量。次级划分变量为坡向（ASP）与土壤容重（BDW）。图 3-8 (b) 使用四种颜色标识出地理最优的四个分区，其特征分别为：高总氮-低坡向、高总氮-高坡向、低总氮-高容重，以及低总氮-低容重。

2021 年中，划分的首要变量为增强型植被指数（EVI），次级变量为坡向（ASP）与粉粒含量（SLT）（图 3-8 (d) 与 (e)）。通过两个层次的变量分层，这三个变量将研究区域划分为四个主要类别。随后在更多变量条件下进一步细化分层，从而在确保四大类之间存在显著空间差异的同时，减少每类内部 LGA 间的差异性。2021 年四类区域的特征分别为：低 EVI-低坡向、低 EVI-高坡向、高 EVI-低粉粒含量，以及高 EVI-高粉粒含量。

图 3-8 (c) 与 (f) 中的箱线图显示，两年份划分所依据的主要变量成功地按照不同地理特征对产区进行了区分。地理最优区域的识别有助于深入理解小麦产量空间异质性在外部决定因素影响下的表现，同时为不同地理区域下的精准农业管理与差异化调控提供科学依据，从而提升产量与经济效益。

图 3-9 与表 3-6 展示了各解释变量在划分地理最优小麦产区时的贡献比例。从图中可以看出，地理、土壤、环境与气候因素均对小麦产量空间分布具有重要影响。2016 年，主导变量为地形和土壤条件，说明当年气候和环境条件变化较小，对产量的影响相对较弱。而在 2021 年，主导因素转变为土壤与环境变量，反映出不同年份决定因素的作用呈动态变化。这种变化趋势可能与气候变化、农业技术进步及环境保护政策的实施有关。例如，2021 年土壤改良与植被修复措施的增强，以及环境政策的落地，均对小麦生产产生了显著影响；此外，降水和气温等气候变量的波动也可能在不同时期对产量产生不同程度的影响。

表 3-6 详细列出了各变量对地理最优区域划分的贡献率。2016 年中，坡向（ASP）、总氮（NTO）和有效持水量（AWC）的贡献率分别为 42.29%、24.51% 和 11.07%，表明地形与土壤属性对产量空间差异的影响最为显著。而在 2021 年，增强型植被指数（EVI）和坡向（ASP）的贡献率最高，分别为 37.25% 和 30.07%，其次为总降水量（TP）与 NDVI。这表明该年份中植被状况、地形及降水条件对小麦产量空间分布具有重要影响。ASP 在两个年份中均表现出关键作用，其通过调节光照、辐射、温度、降水与风速，间接影响小麦的光合作用、授粉过程与营养吸收，从而显著影响小麦产量。

上述结果为理解澳大利亚小麦产量的空间差异性提供了理论支撑，同时也为实现农业生产的区域优化管理与政策制定提供了科学依据。

3.4.4 确定变量交互作用的影响

本研究所构建的 GOZH 模型旨在探讨澳大利亚小麦产量的空间差异性，并在不同的热点区、非热点区及各州范围内分别运行，以评估其空间异质性特征。图 3-10 展示了

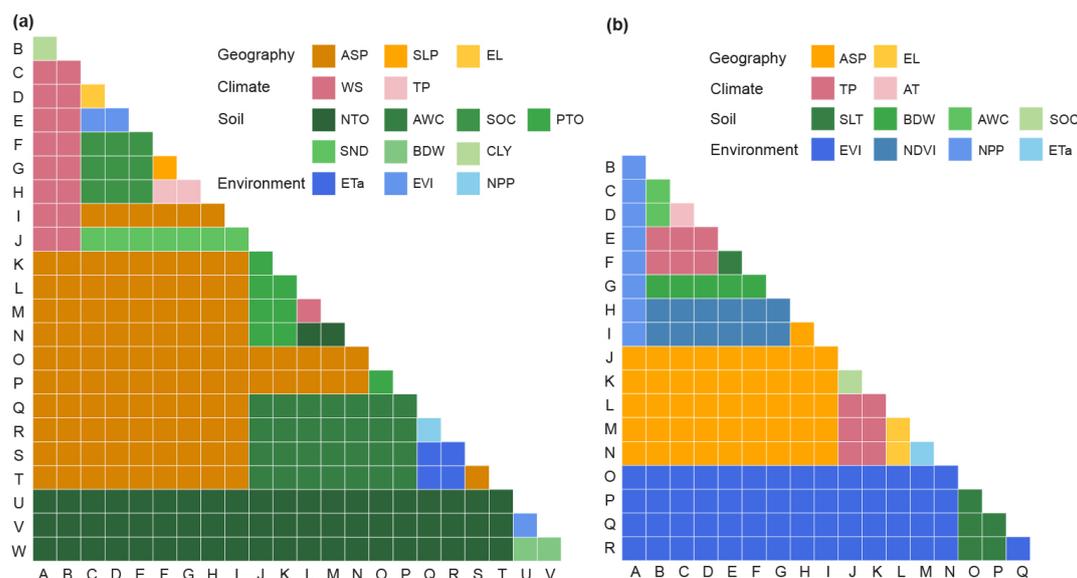


图 3-9 2016 年 (a) 和 2021 年 (b) 解释变量对各对地理最优区域的小麦平均产量的贡献度
 Fig. 3-9 Contributes of explanatory variables for each pair geographical optimal zones of mean wheat production in 2016 (a) and 2021 (b).

表 3-6 解释变量对最佳区域划分的贡献

Tab. 3-6 Contributions of explanatory variables to dividing optimal zones.

| 2016 | | 2021 | |
|----------------|--------------------|----------------|--------------------|
| 变量 Variable | 贡献度 Contributes | 变量 Variable | 贡献度 Contributes |
| ASP | 42.29% | EVI | 37.25% |
| NTO | 24.51% | ASP | 30.07% |
| AWC | 11.07% | NDVI | 7.84% |
| WS | 6.72% | TP | 7.84% |
| SOC | 3.56% | NPP | 5.23% |
| PTO | 3.16% | SLT | 3.92% |
| SND | 2.77% | BDW | 3.27% |
| ETa | 1.58% | AWC | 1.31% |
| EVI | 1.19% | EL | 1.31% |
| TP | 0.79% | ETa | 0.65% |
| BDW | 0.79% | AT | 0.65% |
| NPP | 0.40% | SOC | 0.65% |
| SLP | 0.40% | | |
| EL | 0.40% | | |
| CLY | 0.40% | | |

该模型在澳大利亚多个空间尺度下的运行结果，其中 Ω 值表示地理、气候、土壤和环境因素之间相互作用对小麦产量空间分布的整体影响程度。结果显示，2016 年与 2021 年

GOZH 模型的 Ω 值分别为 0.707 与 0.834，说明交互决定因子组合可以分别解释 70.7% 与 83.4% 的小麦产量空间分布格局。在各热点区与州级尺度上，组合后的 Ω 值均大于 0.5，表明多个因素的交互作用在不同区域均具有显著影响。特别地，2021 年维多利亚州（VIC）与西澳大利亚州（WA）的 Ω 值显著偏高，反映出决定性因子在这些地区的交互影响尤为显著。相比之下，2016 年非热点区域的 Ω 值仅为 0.508，这可能与该区域内小麦产量空间分布较为分散有关。

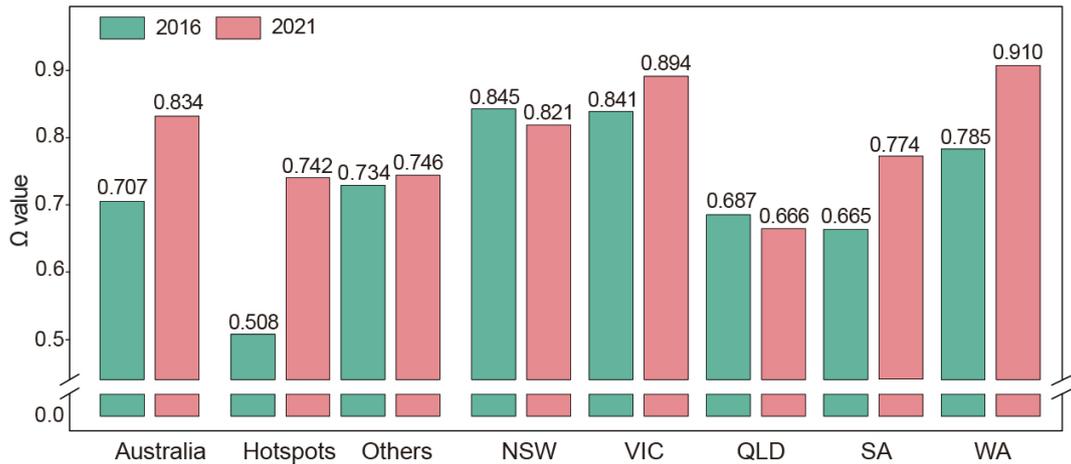


图 3-10 澳大利亚区域、热点与非热点区域及各州基于 GOZH 模型的 Ω 值分布

Fig. 3-10 Ω values in Australia, hotspot and non-hotspot regions, and across individual States using the GOZH model.

这些结果为理解澳大利亚小麦产量的空间差异性及其背后复杂的多因子交互机制提供了重要支撑。研究表明，不同变量在地理空间中的交互作用显著影响小麦产量的空间格局与变异性。

3.4.5 模型验证与评估

本研究进一步通过评估各变量的个体贡献、不同分层水平下解释力（Power of Determinants, PD）值及模型总体 PD 值，全面考察了 GOZH 模型在揭示小麦产量空间差异性方面的性能。此外，为验证 GOZH 模型的有效性，本文将其与另一种地理空间异质性分析方法——基于最优参数的地理探测器模型（Optimal Parameters-based Geographical Detector, OPGD）进行对比分析。

图 3-11 (a) 与 (c) 展示了 OPGD 模型下各解释变量的贡献率，输入变量与响应变量与 GOZH 模型保持一致。结果显示，2016 年 OPGD 模型中，土壤容重（BDW）、有效持水量（AWC）、坡向（ASP）与总氮（NTO）的贡献率分别为 14.2%、9.7%、8.8% 与 8.2%；而 2021 年中，EVI、总降水量（TP）、气温（AT）与粉粒含量（SLT）的贡献率分别达到 36.7%、35.5%、29.5% 与 24.8%。图 3-11 (b) 与 (d) 对比了两个模型在各变量

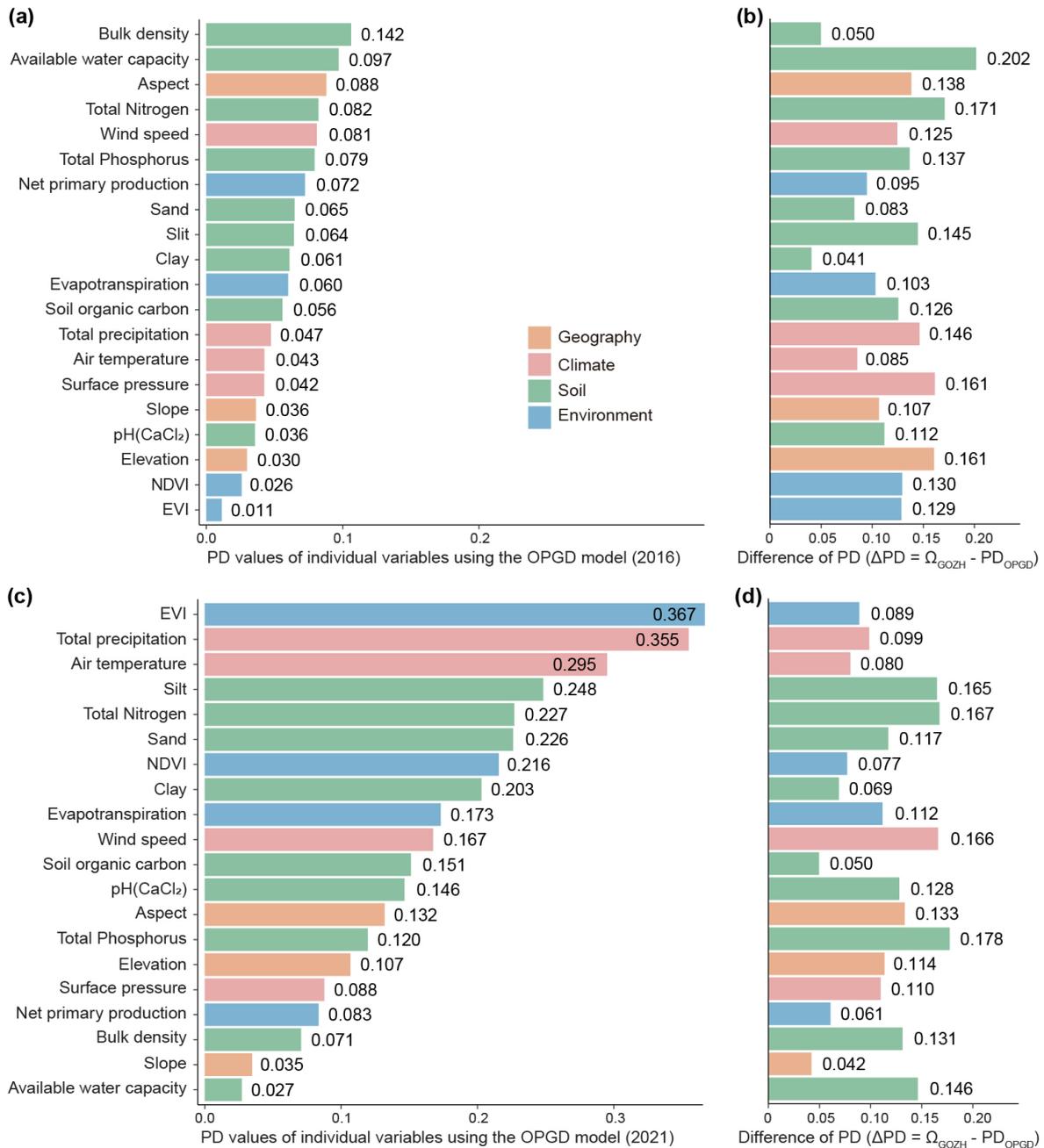


图 3-11 2016 年 (a) 和 2021 年 (c) OPGD 模型中各解释变量的 PD 值, 以及 2016 年 (b) 和 2021 年 (d) OPGD 模型与 GOZH 模型之间各变量 PD 值的差异

Fig. 3-11 PD values of individual explanatory variables in the OPGD model in 2016 (a) and 2021 (c), Differences in individual variable PD values between the OPGD and GOZH models in 2016 (b) and 2021 (d).

贡献率方面的差异。结果表明, 相较于 GOZH 模型, OPGD 模型在变量贡献率上普遍偏低。例如, 2016 年 AWC 在 GOZH 模型中的贡献率比 OPGD 高出 20.2%; 而贡献率差异最小的变量为 CLY。这一结果表明, GOZH 模型中个体变量对小麦产量空间差异性的解释能力更强, 影响更为显著。

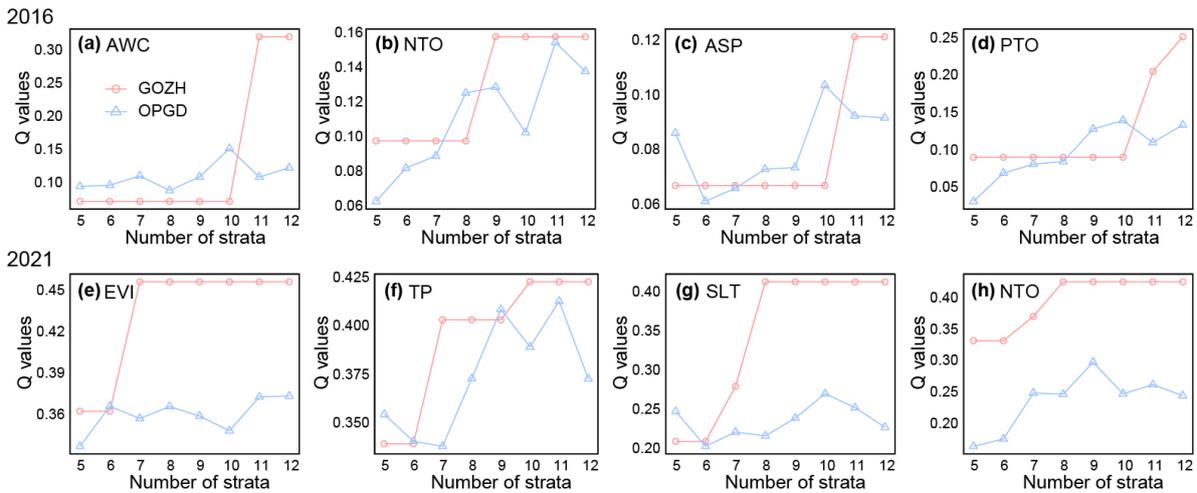


图 3-12 2016 年和 2021 年 GOZH 与 OPGD 模型中四个主要变量在不同分层下的 PD 值变化。2016 年：(a) AWC, (b) NTO, (c) ASP, (d) PTO；2021 年：(e) EVI, (f) TP, (g) SLT, (h) NTO

Fig. 3-12 Variation of PD values for the four principal variables across different stratas in the GOZH and OPGD models in 2016 ((a) AWC, (b) NTO, (c) ASP, (d) PTO) and 2021 ((e) EVI, (f) TP, (g) SLT, (h) NTO).

图 3-12 展示了 2016 年与 2021 年中 GOZH 与 OPGD 模型下，前四个主导变量在不同空间分层水平下 PD 值的变化情况。图 3-12 (a)-(d) 分别对应 2016 年 AWC、NTO、ASP 与 PTO 变量的 PD 值变化；图 3-12 (e)-(h) 展示了 2021 年 EVI、TP、SLT 与 NTO 在不同空间分层下的 PD 值变化趋势。从图中可以明显看出，GOZH 模型在不同变量与层级下的总体 PD 值均高于 OPGD 模型。GOZH 模型的 PD 值随分层数的增加迅速达到最优，而 OPGD 模型的 PD 值则波动较大，整体处于较低水平。以 2021 年为例，EVI 在第七层级达到最优 PD 值，SLT 与 NTO 则分别在第八层级达到最优。这进一步表明，GOZH 模型在不同空间层级下表现出更高的稳定性与解释力。

最终，两个模型在 2016 年与 2021 年的总体 PD 值分别为：2016 年 OPGD 为 0.42，GOZH 为 0.71；2021 年 OPGD 为 0.65，GOZH 为 0.83。结果清晰表明，GOZH 模型在解释澳大利亚小麦产量空间差异性方面具有更优性能，表现出更强的地理异质性识别能力。

3.5 讨论

本研究通过整合空间自相关、空间分层异质性与决策树分析，构建了地理空间机器学习模型，用以识别影响澳大利亚小麦产量空间差异的决定因素。首先，利用 Moran's I 与 LISA 方法对小麦产量的空间自相关进行检验，并结合热点与显著性分析。随后，根据多变量交互作用，将研究区划分为四个地理最优区域 (Geographically Optimal Zones)，从而锁定支配各区产量差异的关键因子。该模型在地理探测器框架基础上，通过逐步

空间离散化与区域最优化进一步量化区域间异质性；GOZH 模型在评估澳大利亚小麦产量空间变异方面表现出较高的精度与可靠性。研究成果有助于揭示小麦产量空间格局，为实施精准农业、改良土壤健康、优化水资源管理及筛选适宜品种等提供决策支持，从而提升产量并保障粮食安全。

澳大利亚幅员辽阔，人口分布稀疏，尤其是中西部地区沙漠与干旱土地广布，土壤类型普遍不利于农业生产。相比之下，东部、东南部及西南部降水和温度适宜，土壤富含有机质，更适合作物生长，致使小麦主要集中于这些区域。气候因子——尤其温度、降水、大气压与风速——对小麦生长至关重要。温带气候（主要分布于新南威尔士州、维多利亚州、南澳大利亚州及西澳大利亚州南部）冬季降水充足，对小麦生长尤为关键；亚热带气候（主要位于昆士兰州东部与新南威尔士州北部）冬季较为温和；地中海气候（西澳大利亚州西南部及部分南澳大利亚州地区）同样提供有利的冬季降水及适中温度。考虑到州际环境差异显著，本研究采用空间自相关方法解析小麦产量的聚集与热点分布，以更充分地理解其地理与空间变异特征。

影响小麦产量的因素十分复杂，其中地形、气候、土壤养分与环境条件被普遍认为是主要驱动 (Ajami et al. 2020; Fan et al. 2018; Hodson and White 2007)。本文选取地理、气候、土壤性质与环境四大类变量作为自变量，通过空间聚类、区域最优化、单变量与多变量交互分析以及模型比对，系统讨论了澳大利亚小麦产量的空间差异，并确定各区域的主导因子。地形与土壤属性直接影响小麦的生境与养分吸收，气候与环境因子则作用于全生育期并调控作物发育；这些因素相互作用，共同塑造小麦最终产量及其空间分布。例如，2020–21 年受拉尼娜湿润条件影响，冠层绿度增强，导致增强型植被指数 (EVI) 在当年成为最重要的决定因子 (Bureau of Meteorology 2022)；此外，新南威尔士州与维多利亚州灌溉效率的提升进一步放大了这一植被信号 (Hughes et al. 2020)。

本研究构建的地理空间机器学习模型融合了机器学习、地理探测器与分层异质性分析，克服了传统模型忽视空间特征且多局限于单因子解析的不足 (Luo et al. 2022)，从多变量交互与地理视角全面探究了小麦产量差异的决定机制，为农业领域空间相关与差异研究提供了新的方法探索。

综上，研究小麦产量的空间异质性对于提高粮食生产效率、保障粮食安全、优化资源管理、制定更有效的农业政策及推动农业科学发展具有重要意义，对全球农业系统的可持续性和未来粮食供给同样至关重要。

3.6 小结

本章构建了一种地理空间机器学习模型，将空间自相关、空间分层异质性与决策树分析相结合，主要使用 GOZH 模型来揭示澳大利亚小麦产量的空间差异及其影响因素。研究表明，小麦产量在多重因素作用下呈现显著空间分异，其中地形、气候、土壤属性与环境条件的交互作用是决定产量差异的最关键因素。基于两个年份的案例分析，模型识别出导致空间差异的主导因子，并将研究区划分为四个地理最优区域，每个区域呈现不同的因子分布特征，为因地制宜的播种、栽培与管理提供了有力依据。

GOZH 模型与 OPGD 模型的对比结果表明，GOZH 在单变量 PD 值上始终更高，且能更快达到最优 PD 值，说明 GOZH 在解释作物产量空间差异方面更具优势。本研究从宏观尺度深化了对澳大利亚小麦产量空间分异及其决定因素的认识。所提出的模型框架可推广至其他作物及地区，用于分析不同作物在多样环境中的空间差异，从而更有效地提升全球粮食生产与保障粮食安全。

第四章 第二维度异常方法的构建及应用

本章提出了第二维度空间异常的概念，并基于该概念开发了一个新的空间预测方法。该方法通过识别样本点一定邻域范围内的解释变量所对应的空间异常点，基于这些正负空间异常构建异常强度指数，并生成解释变量对应的 N 多个第二维度异常变量，这些变量结合原有样本点位置的解释变量，共同构建第二维度异常空间预测模型，实现了对空间数据的准确预测。本章将该方法应用于澳大利亚小麦产量的空间预测，该方法相较于传统的非空间机器学习方法在精度的提高和误差的降低方面都非常显著，验证了该方法在空间数据预测中的有效性。

4.1 概述

长期以来，空间预测方法经历了从经典地统计学模型到现代机器学习算法的演进，逐步构建起多层次的空间建模方法体系。以克里金 (Kriging) 为代表的地统计学方法，在假定变差函数平稳和空间各向同性的前提下，利用协方差结构对未知点进行最优无偏估计，具有坚实的理论基础与不确定性量化能力。然而，该类模型在面对空间非平稳性、异质性或高维多源变量时预测精度下降，难以捕捉复杂地理过程的非线性特征。随后兴起的地理加权回归 (GWR) 及其多尺度扩展 (MGWR) 通过引入空间权重对局部参数进行拟合，在一定程度上缓解了全局模型对空间结构的简化假设，但仍受到变量选择敏感性、维度灾难和共线性的制约。空间误差模型与空间滞后模型等空间计量经济方法则以权重矩阵形式显式表达空间依赖，但对线性结构和正态误差假设要求较高，对局地异常与复杂变量交互的响应能力有限。

随着机器学习的快速发展，随机森林 (RF)、支持向量机 (SVM)、梯度提升树 (XGBoost) 以及深度神经网络等方法广泛应用于遥感反演、土壤属性估算、作物产量预测等空间建模任务中。这些模型通过强大的非线性拟合能力和对高维特征的适应性，显著提升了预测精度，尤其在多变量耦合建模场景中表现优越。然而，大多数机器学习模型未能显式引入空间结构约束，往往通过引入经度纬度、邻域统计量或叠加克里金残差等方式被动补偿空间依赖，难以充分揭示空间异质性强烈区域的局地波动与异常响应。特别是在自然异构景观或极端气候背景下，传统模型往往将样本点以外的空间异常视为噪声，并在空间平滑过程中将其弱化或抹平，从而限制了模型对局地极端的识别与响应能力。针对上述问题，本文提出“第二维度异常” (Second-Dimension Outliers, SDO) 建模框架，通过多尺度缓冲区系统性识别样本点之外的正负异常信息，并构建 SDO 变量，将其纳入机器学习模型中，使模型具备捕捉空间异常信号的能力，从而提

升在复杂空间背景下的预测精度与鲁棒性。

为解决上述问题，研究者逐渐关注如何将空间结构信息纳入机器学习模型中，以提升其对局部变异的响应能力。当前一些方法尝试引入空间滞后项、空间邻接矩阵或空间权重因子，以增强模型对空间依赖的感知，但这些方法仍主要依赖于样本点之间的空间关系，难以充分利用样本点之外的信息。现实地理环境中，一些关键的空间特征和环境扰动往往出现在非采样位置，如极端气候带、边缘地带或局地异常斑块，若忽略这些“第二维度”的空间信息，将导致模型对地理异常模式的识别不足，进而降低预测准确性。因此，在现有空间预测模型的基础上，有必要进一步发展能够捕捉局地空间异常、整合样本点外空间信息的建模方法，为此，本文提出了第二维度异常（Second-Dimension Outlier, SDO）建模框架，以期突破当前空间预测模型在空间异常响应能力方面的局限，提升复杂空间背景下的预测稳定性与泛化能力。

4.2 第二维度空间异常方法

4.2.1 第二维度空间异常的概念

现有的空间预测模型主要依赖于来自样本点的信息，通常通过机器学习或回归方法预测响应变量的空间分布。然而，这类方法往往忽视了样本点周边未知区域或非采样点位置的数据，这些信息对于提高响应变量的预测精度具有潜在价值(Song 2022)。为突破该限制，本文提出一种新颖的空间建模方法——第二维度异常值模型（SDO）。SDO方法通过在每个样本点周围定义局部缓冲区，识别其邻域内的空间异常值，从而提供额外的空间上下文信息。随后，基于这些邻近点的异常值信息生成一组 SDO 变量，并与原始样本点的解释变量相结合，用于构建更精确的预测模型。

与传统非空间模型相比，SDO 方法具备多个显著优势。首先，引入局部异常信息可以更有效地刻画空间异质性，从而弥补常规模型在该方面的缺失。其次，利用异常值可以提升预测精度，改善最大值与最小值的拟合效果，降低整体预测误差。此外，SDO 模型可用于数据的降尺度处理，使得从粗分辨率的观测数据中实现高精度预测，生成更高空间分辨率和更具细节的预测图层，这对于环境监测与农业预报具有重要意义。

4.2.2 SDO 模型

本研究基于 SDO 的概念开发了 SDO 模型，旨在通过量化局部异常值强度，建立响应变量与空间异常信息之间的关联。设已知样本点位置为 u ，未知网格点为 v 。在本研究中，未知位置被定义为响应变量采样点之外的空间位置与区域。

SDO 模型的构建包含四个主要步骤（见图 4-1）：

第一步是确定一组局部范围（即缓冲区大小），用于在不同空间尺度下计算解释变量的异常值信息。缓冲区大小的选择基于研究区域响应变量的空间特征、尺度与属性。

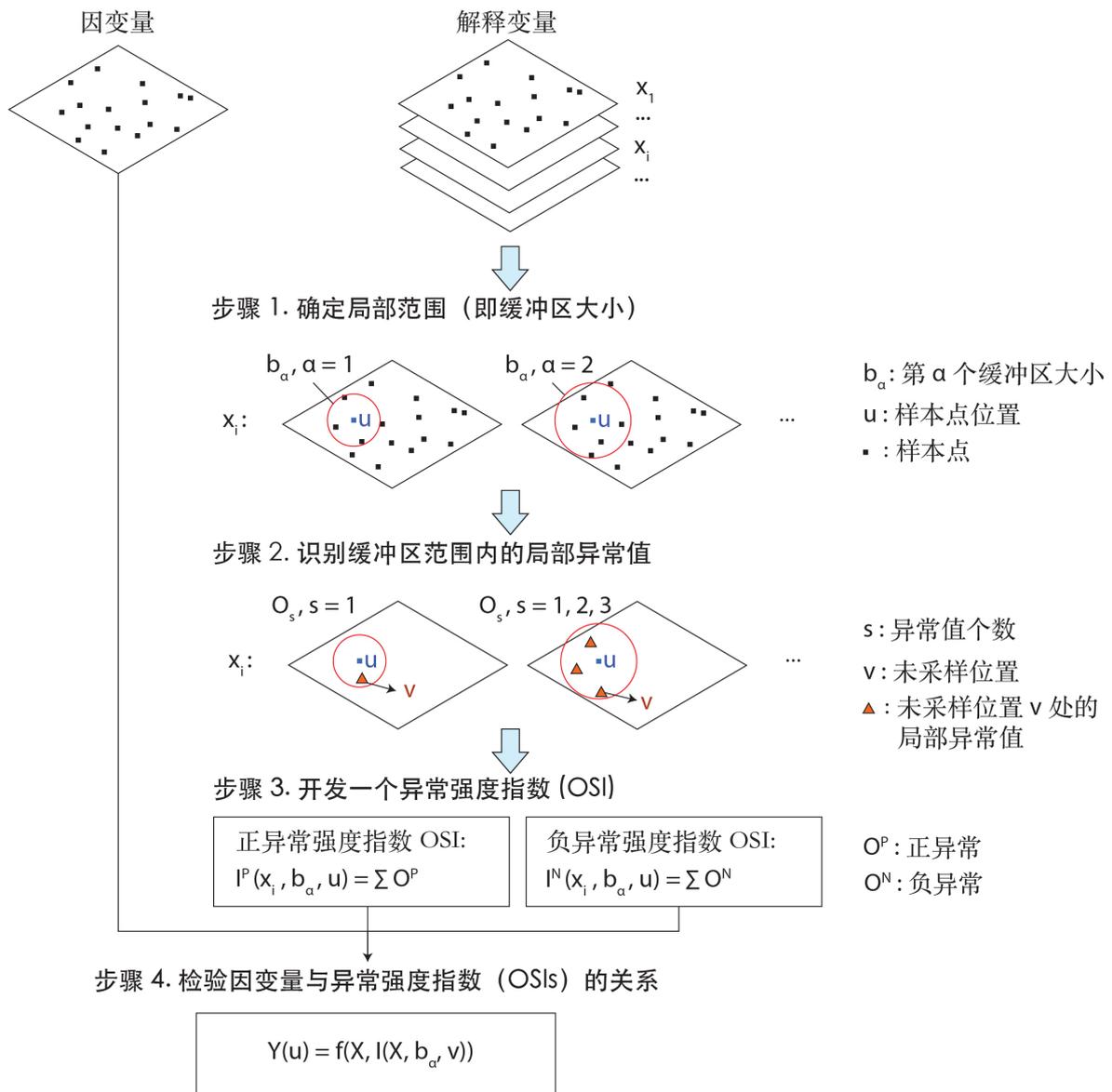


图 4-1 用于分析因变量与空间局部离群点关系的二阶离群点 (SDO) 模型示意图。

Fig. 4-1 Schematic overview of second-dimension outliers (SDO) model for examining the relationship between dependent variables and spatial local outliers.

较大的缓冲区可捕捉更多空间异常信息，但也可能引入过多外围影响，从而干扰中心采样点的估计，降低整体预测精度。因此，缓冲区大小应与研究区的空间尺度相匹配。

一般而言，缓冲区数量应控制在 5 至 10 个范围内，以平衡计算效率与统计可靠性，避免样本量过少或数据量过大。缓冲区阈值通常设定为样本点间最大两两距离的 10% 至 20%，以确保在适当空间范围内获取异常值信息(Qi et al. 2022; Shi et al. 2021; Zhang et al. 2021)。根据上述标准，本研究在模拟数据集中采用了 6 个缓冲尺度 (范围 2 至 7)，在实证案例数据集中采用了 7 个缓冲尺度 (范围为 1° 至 7°)。缓冲区阈值与间隔的计算公式如下：

$$b_\alpha, \alpha = 1, 2, \dots, m \quad (4-1)$$

式中, b_α 表示第 α 个缓冲区的大小, m 为缓冲阈值, K 为缓冲区个数。

第二步是识别各缓冲区范围内的局部异常值。对于每个缓冲尺度, 通过比较样本点周围缓冲区内点值与中心点值, 识别出其正向与负向异常值。异常值依据原有样本点数据相对于其期望值(均值)的偏离程度来定义, 若值大于 $\bar{x} + 2\sigma$, 则视为正向异常值 (O_s^P); 若小于 $\bar{x} - 2\sigma$, 则为负向异常值 (O_s^N)。对于某一解释变量, 其在位置 u 的局部异常值定义如下:

$$O_s(u), s = 1, 2, \dots, n \quad (4-2)$$

式中, $O_s(u)$ 为在缓冲区内未知位置 v 的异常值向量, s 表示第 s 个异常值, n 为缓冲区内异常值总数。

第三步是构建异常强度指数 (Outlier Strength Index, OSI), 用于量化每个解释变量在不同尺度下的正向或负向异常程度。OSI 通过对多个缓冲尺度下的所有异常值进行累加计算, 确保各尺度下的空间异常信息均被有效整合。其计算公式如下:

$$\begin{aligned} I^P(X, b_\alpha, v) &= \sum_{s=1}^m O_s^P(b_\alpha, v) \\ I^N(X, b_\alpha, v) &= \sum_{t=1}^n O_t^N(b_\alpha, v) \\ I(X, b_\alpha, v) &= [I^P(X, b_\alpha, v), I^N(X, b_\alpha, v)] \end{aligned} \quad (4-3)$$

式中, X 表示空间解释变量, b_α 是第 α 个缓冲距离, v 表示位于目标样本位置 u 周围缓冲区 b_α 内的一个未采样位置。 $O_s^P(b_\alpha, v)$ 和 $O_t^N(b_\alpha, v)$ 分别表示在该缓冲区内基于 z-score 阈值识别出的正向和负向异常值。 I^P 和 I^N 分别表示在特定尺度下的累计正向与负向异常强度指数。每一个缓冲区对于每个解释变量都会产生一组 I^P 和 I^N 值, 从而构成一组 SDO 变量。这些变量被作为多尺度空间特征引入预测模型中, 使模型能够考虑不同距离尺度下的空间异常现象, 从而提升预测精度。

第四步是将原始样本点的解释变量与 SDO 变量 (即 I^P 与 I^N) 结合, 构建空间预测模型。采用随机森林 (Random Forest, RF) 与支持向量机 (Support Vector Machine, SVM) 等机器学习算法进行建模, 并对比 SDO 模型与传统非空间模型在预测精度方面的差异。本文最终构建的空间预测模型的基本形式如下:

$$Y(u) = f(X, I(X, b_\alpha, v)) \quad (4-4)$$

式中， $Y(u)$ 表示在采样位置 u 处的空间响应变量。 X 表示原始的解释变量，在采样点位置上被观测。 $I(X, b_\alpha, v)$ 表示第二维度异常变量，是在不同缓冲区大小 b_α 内，基于未采样位置 v 处的 X 的异常强度所构建的变量。函数 $f(\cdot)$ 表示用于空间预测的模型（例如支持向量机 SVM、随机森林 RF），该模型同时利用原始解释变量 X 以及所提取的异常变量进行训练。需要特别指出的是，该方法并未从 X 中剔除异常值，而是通过提取和量化异常信息，以增强空间预测能力。

4.3 材料与方方法

4.3.1 数据来源及预处理

本章将第二维度异常值模型（SDO）应用于澳大利亚小麦产量的空间预测，重点关注 2021 年澳大利亚小麦带内 179 个地方政府区域（Local Government Areas, LGAs）(Feng et al. 2022)。小麦产量数据来源于澳大利亚统计局（ABS）(Australian Bureau of Statistics 2021b)，以每个 LGA 的平均产量作为预测模型的响应变量。图 4-2 展示了 2021 年澳大利亚小麦带内 179 个 LGA 的小麦平均产量分布，其中主要产区集中在新南威尔士州（NSW）中部、维多利亚州南部和西澳州西部。

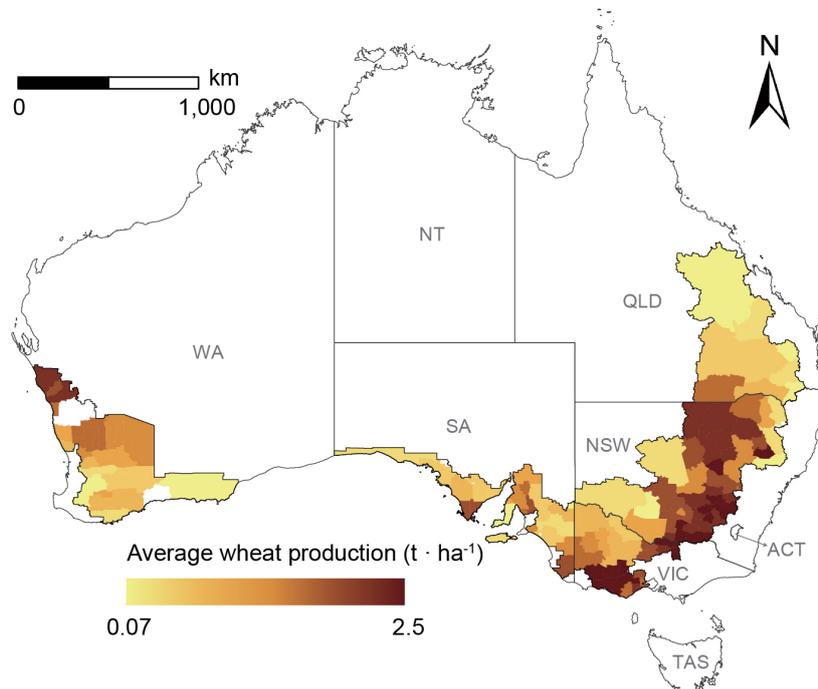


图 4-2 2021 年澳大利亚小麦带内各 LGA 的年度小麦产量。

Fig. 4-2 Annal wheat production in Australia's LGAs within the wheat belt in 2021.

关于影响小麦产量的解释变量，本章根据第三章的分析结果以及文献资料，选取

了三大类具有显著作用的数据：气候因子、环境因子与土壤属性，具体变量见表 4-1。气候变量包括气温与总降水量，数据源自 ERA5-Land 再分析数据集，提供了 2021 年覆盖小麦带的 0.25° 分辨率气候平均值 (Hersbach et al. 2023)。环境变量包括实际蒸散 (ETa)、归一化植被指数 (NDVI) 和增强型植被指数 (EVI)。ETa 数据来自 Google Earth Engine 平台上的 CMRSET Landsat V2.2 数据集，空间分辨率为 30 米 (Guerschman et al. 2022)；NDVI 与 EVI 数据则来自 MOD13A2 V6.1 产品，分辨率为 1000 米 (Didan 2021)。土壤数据来自 CSIRO 发布的澳大利亚土壤与地貌网格 (SLGA) 数据集，选取了总氮 (NTO)、总磷 (PTO)、砂粒比例 (SND) 与粉粒比例 (SLT) 等对小麦生长关键的土壤属性 (Rossel et al. 2015)。图 4-3 展示了小麦带内这些解释变量的空间分布情况。

表 4-1 潜在影响小麦产量空间差异的解释变量概述。

Tab. 4-1 A summary of explanatory variables that potentially affect spatial disparities of wheat production.

| 类别 Category | 变量 Variable | 代码 Code | 数据产品 Product | 分辨率 Resolution |
|-------------------|---|------------|---------------------|-------------------|
| 气候 Climate | 气温 Air temperature | AT | ERA5_land | 0.25° |
| | 总降水量 Total precipitation | TP | ERA5_land | 0.25° |
| 环境 Environment | 蒸散量 Evapotranspiration | ETa | CMRSET Landsat V2.2 | 30 m |
| | 归一化植被指数 Normalized difference vegetation index | NDVI | MOD13A2 V6.1 | 1000 m |
| | 增强型植被指数 Enhanced vegetation index | EVI | MOD13A2 V6.1 | 1000 m |
| 土壤数据 Soil data | 总氮 Total Nitrogen | NTO | CSIRO/SLGA | 92.77 m |
| | 总磷 Total Phosphorus | PTO | CSIRO/SLGA | 92.77 m |
| | 砂粒 Sand | SND | CSIRO/SLGA | 92.77 m |
| | 粉粒 Silt | SLT | CSIRO/SLGA | 92.77 m |

由于本研究仅考虑位于澳大利亚小麦带内的地方政府区域 (LGA)，首先计算了这些 LGA 的平均小麦产量，并在进行平均处理前，利用全国小麦种植图对所有环境变量图层进行了掩膜操作，确保仅保留小麦种植区域内的像素值。在预测格网的处理中，每个解释变量图层均在 Google Earth Engine 平台上进行了重采样，使得每个格网单元内的平均值能够代表该变量在更精细空间尺度上的特征。上述步骤有效减少了生态尺度偏差，同时确保响应变量与其协变量之间具有内部一致性。

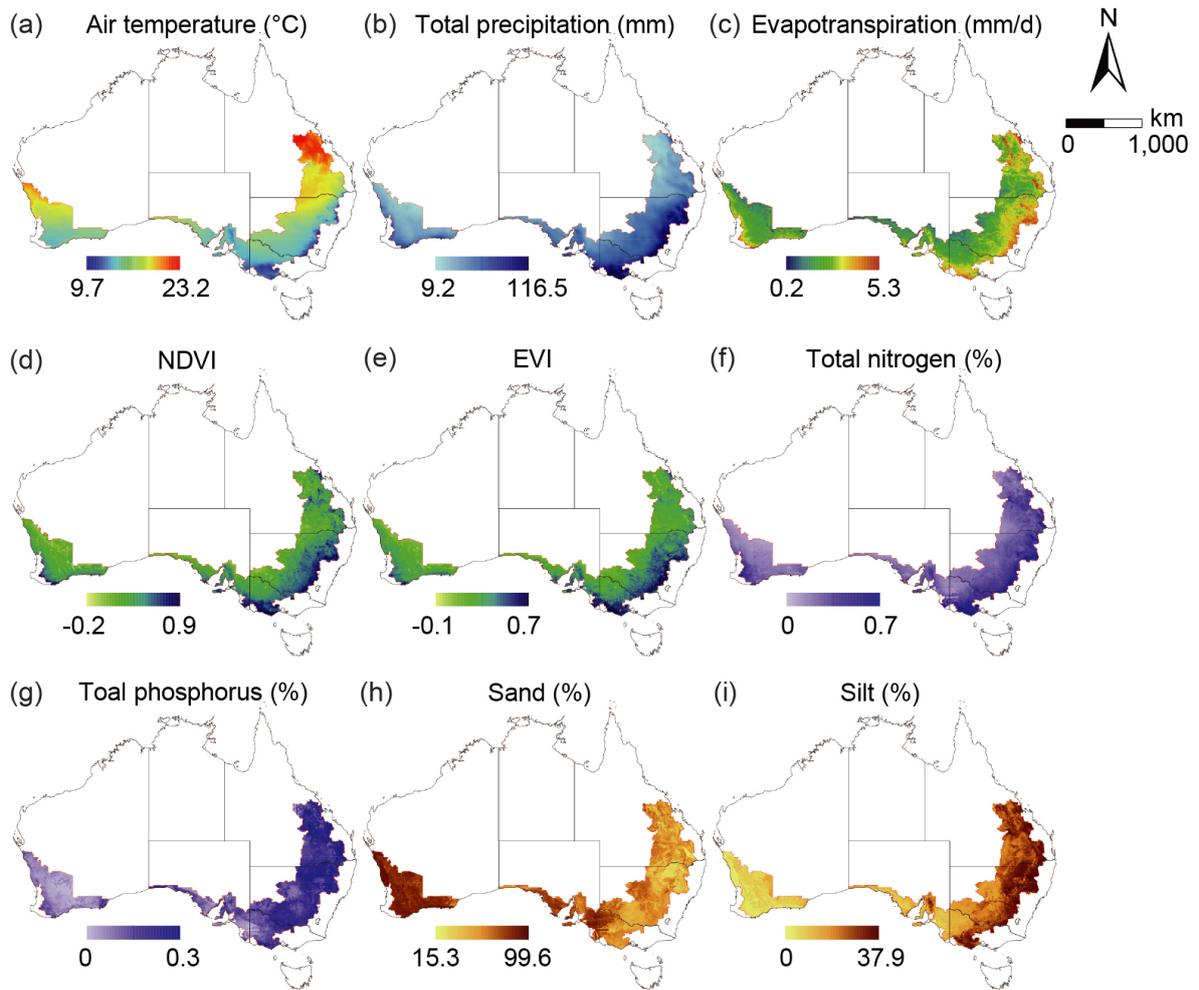


图 4-3 2021 年解释变量的空间分布。(a) - (b) 气候变量, (c) - (e) 环境变量, (f) - (i) 土壤属性。
 Fig. 4-3 Spatial distributions of explanatory variables in 2021. (a) - (b) Climate variables, (c) - (e) Environmental variables, (f) - (i) Soil attributes.

4.3.2 基于 SDO 模型的模拟数据

本研究首先通过模拟数据集评估所提出的第二维度异常值模型 (SDO) 的性能。在模拟实验中, 于一个 20×20 的二维空间内随机生成了样本点的经纬度位置。响应变量 y 以及四个解释变量 x_1 、 x_2 、 x_3 和 x_4 亦为随机生成。图 4-4 展示了响应变量 y 与四个解释变量的空间分布情况, 以及变量间的相关性分析结果。

模拟样本数据集包含经度 (lon)、纬度 (lat)、响应变量 y 及四个解释变量 (x_1 、 x_2 、 x_3 、 x_4)。构建的非空间预测模型 (aspatial model) 旨在以这四个解释变量为输入预测 y 的值。整个数据集中共包含 300 个样本点, 其中 y 的取值范围为 3.42 至 8.26, 均值为 6.09。预测区域采用标准 20×20 的规则网格, 经纬度范围均为 1 至 20, 每个网格点均包含相同的四个解释变量值 (x_1 至 x_4), 为模型提供统一的空间预测框架。在模型评估中, 采用五折交叉验证 (five-fold cross-validation) 方法, 将数据集划分为 80%

(240 个样本点) 用于训练, 20% (60 个样本点) 用于验证, 以确保模型评估的稳健性。

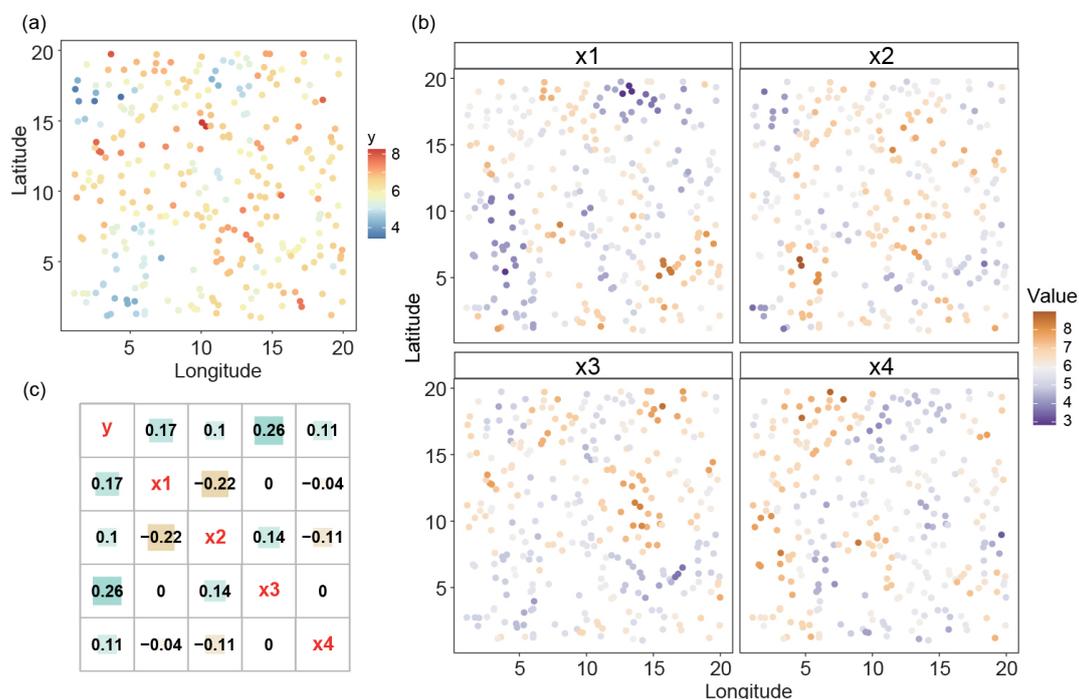


图 4-4 模拟数据的空间分布情况, 包括因变量 (a)、解释变量 (b) 及其相关性分析 (c)。

Fig. 4-4 Summary of simulation data including dependent variable (a), explanatory variables (b), and the correlation analysis with them (c).

4.3.3 随机森林回归方法 (RF)

设训练样本为 $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, 其中 $\mathbf{x}_i \in \mathbb{R}^d$ 为 d 维特征向量 (可包含经纬度或派生的空间协变量), $y_i \in \mathbb{R}$ 为响应变量 (产量、环境指标等)。随机森林通过以下步骤进行空间回归预测:

1) 树模型构建与集成 对 $t = 1, \dots, T$:

1) 采用自助抽样从原始数据生成子集 \mathcal{D}_t ;

2) 在 \mathcal{D}_t 上训练回归树 f_t , 在每次节点分裂时随机选取 m_{try} 个特征并以平方误差最小化准则确定最优划分;

集成预测为

$$\hat{f}(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^T f_t(\mathbf{x}). \quad (4-5)$$

2) **袋外误差与变量重要性** 对每个样本,统计未被抽入 \mathcal{D}_i 的 OOB 树预测平均值 $\hat{f}_{\text{OOB}}(\mathbf{x}_i)$, 计算

$$\text{OOB-RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{f}_{\text{OOB}}(\mathbf{x}_i))^2}.$$

对变量 x_j , 在各 OOB 树中随机置乱 x_j 并记录误差增加量, 平均后即为其 **置换重要性**。

3) **空间特征工程与预测栅格化** 空间预测通常面临自相关和尺度异质性。常用策略包括:

- **显式坐标法**: 将经纬度或空间平滑基函数 $\{\phi_k(\text{lon}, \text{lat})\}$ 纳入 \mathbf{x} 中, 使模型显式学习空间趋势;
- **邻域统计**: 对栅格或点数据计算邻域平均、方差或自相关系数作为派生特征;
- **残差-克里金法 (RF + OK)**: 先用 RF 拟合非空间协变量, 得到残差 $e_i = y_i - \hat{f}(\mathbf{x}_i)$; 再对 e_i 进行普通克里金 (OK) 插值, 最终预测

$$\hat{y}(\mathbf{s}) = \hat{f}(\mathbf{x}(\mathbf{s})) + \hat{e}_{\text{OK}}(\mathbf{s}). \quad (4-6)$$

4) **空间交叉验证** 为避免空间自相关导致的乐观偏差, 采用基于分区或缓冲的 *spatial block cross-validation*: 在训练/验证拆分时确保测试单元在地理上与训练单元分离, 以获得更真实的泛化误差估计。

5) **不确定性量化** 可利用森林中各树的预测分布近似 $\hat{f}(\mathbf{x})$ 的置信区间:

$$\hat{\sigma}^2(\mathbf{x}) = \frac{1}{T-1} \sum_{t=1}^T (f_t(\mathbf{x}) - \hat{f}(\mathbf{x}))^2,$$

或结合无穷小刀切法 (Infinitesimal jackknife) (Brokamp et al. 2017) 生成更稳健的置信区间, 为空间规划提供风险评估。

随机森林在空间预测中的成功依赖于高质量的空间协变量构建与合理的空间交叉验证; 通过式 (4-5)-式 (4-6) 的集成与残差插值框架, RF 可在捕获复杂非线性关系的同时补偿空间自相关特征, 广泛用于作物产量估算、土壤属性制图、森林生物量反演及环境风险评估等场景。

4.3.4 支持向量机回归方法 (SVM)

设训练样本为 $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, 其中 $\mathbf{x}_i \in \mathbb{R}^d$ 为 d 维特征向量, $y_i \in \mathbb{R}$ 为实值响应。在线性可分场景下, SVM 旨在学习函数

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b, \quad (4-7)$$

使得大部分样本落入宽度为 2ε 的管道内；对管外样本以松弛变量 (ξ_i, ξ_i^*) 度量偏离程度。其原始优化问题写作

$$\begin{aligned} \min_{\mathbf{w}, b, \xi_i, \xi_i^*} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & \begin{cases} y_i - \mathbf{w}^\top \mathbf{x}_i - b \leq \varepsilon + \xi_i^*, \\ \mathbf{w}^\top \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i, \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, N, \end{cases} \end{aligned} \quad (4-8)$$

式中 $\frac{1}{2} \|\mathbf{w}\|^2$ 为模型复杂度项， $C > 0$ 为惩罚系数，用于平衡模型平滑性与对管外误差的容忍度。通过拉格朗日乘子法可将问题 (式 (4-8)) 转化为对偶形式

$$\begin{aligned} \max_{\alpha_i, \alpha_i^*} \quad & -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \\ \text{s.t.} \quad & \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0, \quad 0 \leq \alpha_i, \alpha_i^* \leq C, \end{aligned} \quad (4-9)$$

式中 $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$ 为核函数， $\phi(\cdot)$ 为隐式高维映射。对偶问题 (式 (4-9)) 仅涉及样本内积，可通过核函数实现非线性回归而无需显式计算高维特征。最优解 (α_i, α_i^*) 满足 KKT 条件 (Karush-Kuhn-Tucker Conditions)，通常仅有极少数样本对应 $\alpha_i - \alpha_i^* \neq 0$ ，这些样本即为支持向量。得到乘子后，预测函数表示为：

$$f(\mathbf{x}) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b, \quad (4-10)$$

式中截距 b 可由 KKT 条件或样本平均法求得。常用核函数及其表达式列于表 4-2。核参数 (如 RBF 核的 γ) 与 C 、 ε 共同决定模型的泛化能力，可通过交叉验证或网格搜索进行优化。

表 4-2 支持向量机 (SVM) 常用核函数与主要参数说明

Tab. 4-2 Common Support Vector Machine (SVM) Kernel Functions and Key Parameter Descriptions

| 核类型 | 数学表达式 $K(\mathbf{x}_i, \mathbf{x}_j)$ | 关键参数及含义 |
|------------------|---|---|
| 线性核 | $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j$ | 无 |
| 多项式核 | $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^\top \mathbf{x}_j + r)^d$ | γ : 缩放系数; r : 偏移常数; d : 多项式阶数 |
| Gaussian 核 (RBF) | $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\ \mathbf{x}_i - \mathbf{x}_j\ ^2}{2\sigma^2}\right)$ | σ : 核宽度 ($\gamma = 1/(2\sigma^2)$) |
| Sigmoid 核 | $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^\top \mathbf{x}_j + r)$ | γ : 缩放系数; r : 偏移常数 |

SVM 的优势主要体现在：(1) 以凸二次规划为求解框架，保证全局最优；(2) 通过 ϵ -不敏感损失函数抑制异常点影响，提高对噪声数据的鲁棒性；(3) 解由少量支持向量决定，计算与存储成本与数据维度而非样本量线性相关；(4) 利用核函数在高维特征空间实现复杂非线性映射，无需显式特征工程。得益于上述特性，SVM 在小样本、高维度、非线性且噪声较大的预测任务中表现尤为突出，已成为地理学、农业遥感、金融量化及环境工程等领域的常用基准模型。

4.3.5 SDO 模型预测小麦产量的实验设计

SDO 模型的在小麦产量中的空间预测实验设计如图 4-5 所示。第一步，收集研究区的小麦产量数据与解释变量，包括矢量边界数据、遥感产品及统计数据，并进行数据预处理、变量相关性分析与变量筛选。第二步，在不同缓冲尺度下识别空间异常值，生成对应的第二维度异常变量。第三步，将 SDO 变量与支持向量机 (SVM) 集成进行建模预测。第四步，将 SDO 模型与六种其他机器学习模型结合进行比较分析，以评估 SDO 模型的效果。最后，选取综合性能最优的 SVM 模型进行最终预测，并比较 SDO-SVM 模型与非空间 SVM 模型在变量重要性、NSW 剖面预测表现及各州预测值分布等方面的差异。

第一步中，通过变量相关性分析进行数据预处理和变量筛选，识别出对小麦产量影响显著的解释变量。最终筛选出 9 个相关性较高的解释变量，LGA 的平均值被用于训练样本，而用于预测的网格单元分辨率设为 1000 米。

第二步中，根据 SDO 方法，在七个缓冲距离（从 1° 至 7° ，间隔 1° ）下识别局部空间异常值，并计算异常强度指数 (OSI) 以生成对应的 SDO 变量。根据异常值的符号，SDO 变量被划分为正向 SDO 与负向 SDO 变量。每个变量在七个缓冲区尺度下共生成 14 个 SDO 变量（7 个正向，7 个负向）。

第三步，采用基于 SDO 的机器学习方法对小麦产量进行空间预测。SVM 因其良好的泛化能力、对异常值的鲁棒性与较强的可解释性，被用于构建集成模型 (Cherkassky and Ma 2004; Yu and Kim 2012; Zhu, Hao, et al. 2024)。SVM 通过最小化预测误差，并保持样本点在超平面间距之内，从而实现高效回归 (Yu and Kim 2012)。其核函数将输入映射至高维空间，使得模型可在非线性空间中构建最优回归平面 (Abakar and Yu 2014)。本研究采用高斯径向基函数 (Radial Basis Function, RBF) 作为核函数 (Ding et al. 2021; Wang et al. 2004)，表达式如下：

$$K(\vec{x}_i, \vec{y}_j) = e^{-\gamma \|\vec{x}_i - \vec{y}_j\|^2} \quad (4-11)$$

式中， \vec{x}_i 与 \vec{y}_j 为样本向量， γ 控制样本对回归超平面的影响程度。当 γ 较小时，单个样本对整体模型影响较广，可能成为支持向量；而 γ 较大时，核函数更尖锐，模型复

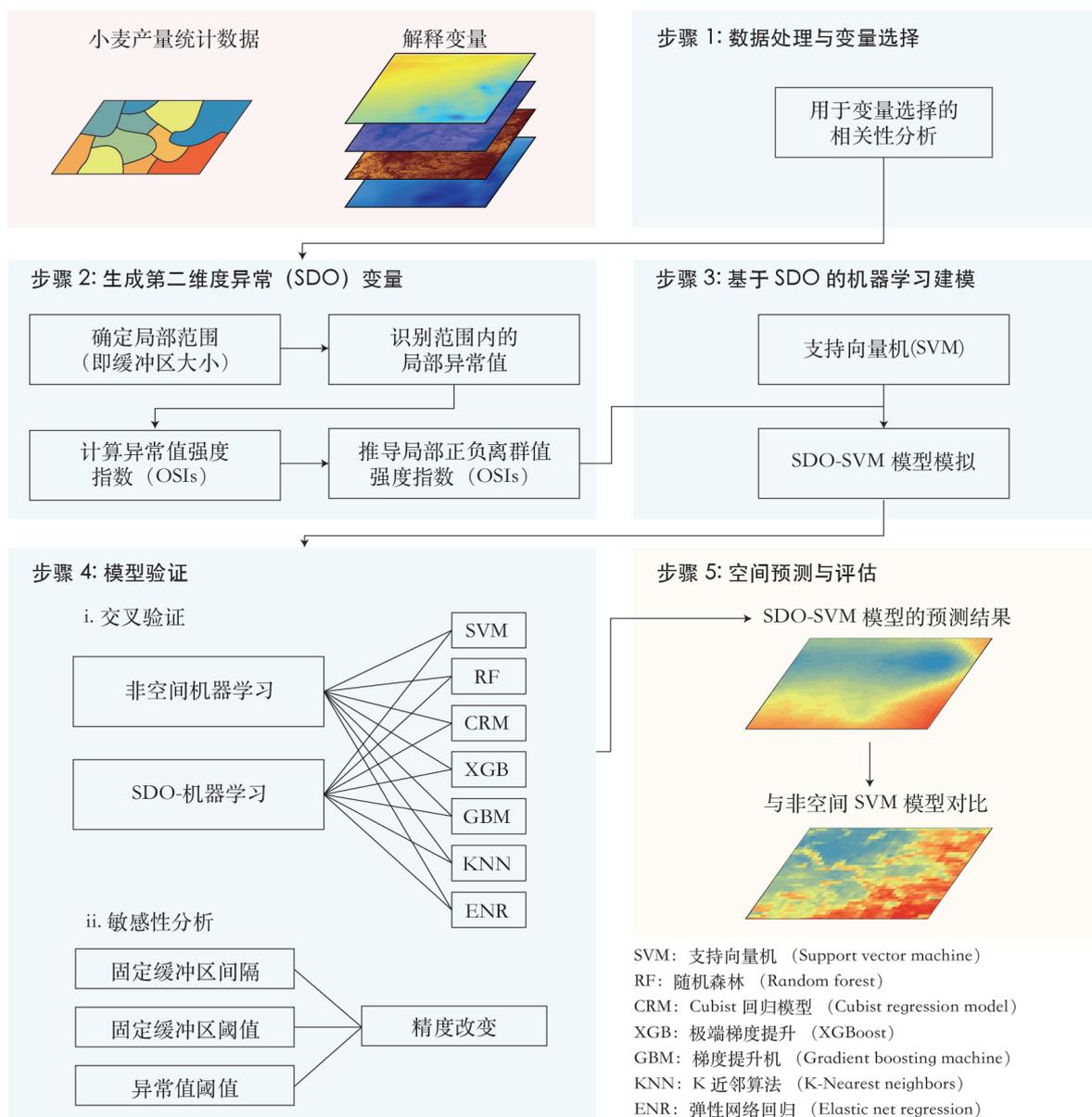


图 4-5 用于预测澳大利亚小麦产量的二阶离群点 (SDO) 模型主要步骤。

Fig. 4-5 Main steps of the second-dimension outliers (SDO) model for predicting wheat production in Australia.

杂度增加, 易发生过拟合。SDO 与 SVM 的集成提升了模型对局部空间异常的鲁棒性, 并优化了超平面的选择, 从而提高回归预测的准确性。

第四步, 对 SDO 模型的预测结果与非空间模型进行比较, 开展敏感性分析以验证模型的稳定性与准确性。在第三步的 SVM 模型基础上, 另选取六种常见机器学习模型进行对比分析, 分别为: 随机森林 (RF)、Cubist 回归模型 (CRM)、极端梯度提升 (XGBoost, XGB)、梯度提升机 (GBM)、K 近邻算法 (KNN) 与弹性网回归 (ENR)。每种算法在非空间数据条件下构建 aspatial 模型 (如 aspatial SVM、aspatial RF、aspatial

CRM 等), 在 SDO 变量参与下构建空间模型 (如 SDO-SVM、SDO-RF、SDO-CRM 等)。所有模型均采用五折交叉验证 (Five-fold Cross-Validation) 进行训练和验证, 每轮使用 80% 数据训练, 20% 用于验证。此外, 还开展了缓冲区间与阈值设置下的敏感性分析, 以测试模型在不同空间参数配置下的稳健性。

预测模型的准确性与误差通过决定系数 (R^2)、均方根误差 ($RMSE$) 与平均绝对误差 (MAE) 进行评估。三种指标的计算公式如下:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4-12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (4-13)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (4-14)$$

式中, y_i 表示观测值, \hat{y}_i 表示预测值, \bar{y} 为观测值均值, n 为样本数量。

最后, 本文基于 SDO-SVM 模型对澳大利亚小麦带区域的小麦产量进行了空间预测, 并对预测结果进行了评估。选取前述表现最优的 SVM 模型与 SDO 集成, 对比其与 *aspatial* SVM 模型的预测性能, 比较两个模型在 R^2 、 $RMSE$ 与 MAE 等指标下的表现。进一步, 分析了新南威尔士州 (NSW) 小麦带区域的剖面预测图, 并比较了各州预测结果的密度分布特征, 进一步验证了本研究提出的 SDO 模型在实际应用中的准确性与有效性。

4.3.6 SDO 模型敏感性分析

为了进一步增强 SDO 模型的鲁棒性, 对其核心参数进行了敏感性分析, 主要关注缓冲区大小、阈值的选择以及异常值的识别阈值对模型预测结果的影响。缓冲区大小的选择直接影响异常值的识别范围, 而缓冲区阈值则决定了哪些异常值被视为正向或负向异常, 异常值的识别阈值即通过标准差来决定识别的空间异常的强度。通过调整这两三个参数, 本研究评估了模型在不同空间尺度下的稳定性与预测精度。

具体的操作为, 缓冲区大小从选择三个范围, 即 6° 、 7° 和 8° 进行变化, 分别计算每个缓冲区下的正向与负向异常值, 并生成对应的 SDO 变量, 结合机器学习模型计算其空间预测的交叉验证均方根误差 (CV-RMSE)。同理, 调整缓冲区阈值 (如 1° 、 2° 和 3°) 来识别正向与负向异常值, 并观察模型在不同参数设置下的预测性能变化。最后, 在异常值识别的阈值的设置上选择常用的 1.5 倍标准差, 2 倍标准差, 2.5 倍标准

差，并比较最终模型 CV-RMSE 的变化情况。通过这种方式，能验证 SDO 模型具有很强的鲁棒性，其对参数的变化敏感性较小，后续可以根据经验和数据分布情况，自行设置更为合适的参数。

4.4 结果分析

4.4.1 SDO 模型模拟数据分析

对于 SDO 模型，首先依据样本点与目标网格点的位置，识别在不同缓冲尺度下的解释变量的正向与负向空间异常值。随后，基于这些异常值生成异常强度指数 (Outlier Strength Index, OSI)，并据此构建对应的 SDO 变量。图 4-6 展示了基于样本点空间位置，对解释变量 x_1 、 x_2 、 x_3 与 x_4 在缓冲尺度 2 至 7 下提取的正向与负向 SDO 变量 (共 24 个)。这些 SDO 变量与原始解释变量一同用于构建预测模型。本研究采用随机森林 (Random Forest, RF) 与支持向量机 (Support Vector Machine, SVM) 等机器学习算法，分别构建非空间模型 (aspatial RF 和 aspatial SVM) 与集成 SDO 变量的模型 (SDO-RF 与 SDO-SVM)，并对预测结果进行比较分析。

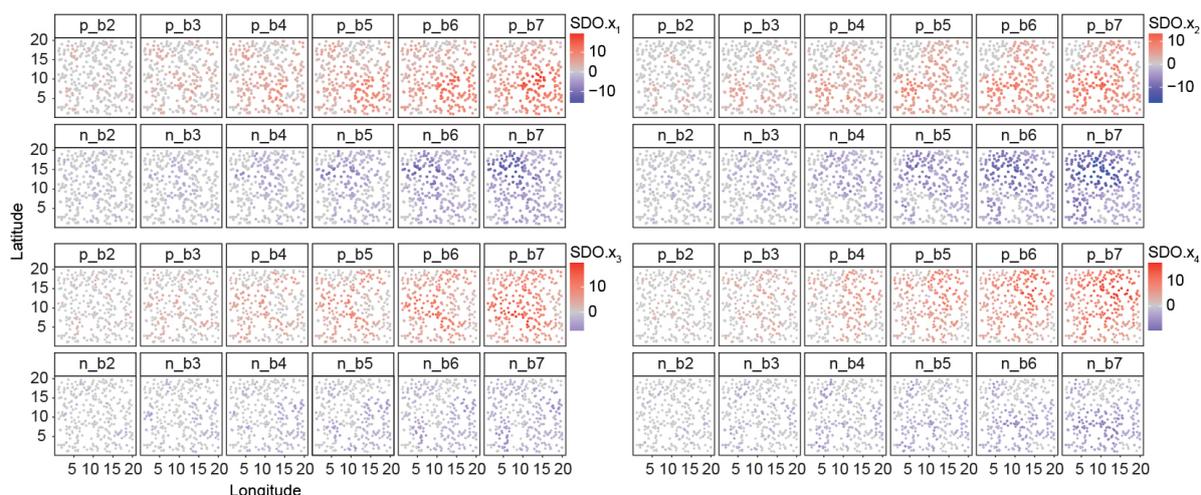


图 4-6 由 SDO 模型生成的模拟数据中解释变量的第二维度变量。其中，字母 p 和 n 分别表示正离群点和负离群点，字母 b_2 、 b_3 、...、 b_7 表示缓冲区大小从 2 到 7 变化的范围。

Fig. 4-6 The second-dimension variables of the explanatory variables of the simulated data are generated by the SDO model. The letters p and n mean positive and negative outliers, respectively. The letters b_2 、 b_3 、...、 b_7 mean the buffers with sizes varying from 2 to 7.

图 4-7 (a) 和 (b) 展示了上述四种模型的空间预测结果，并列出了预测值与真实值之间的皮尔逊相关系数 (R) 作为性能评估指标。结果表明，SDO 模型在预测精度上优于传统非空间模型，预测结果更为平滑，能更准确地捕捉极值区域 (高值和低值)。其中，SDO-SVM 模型的 R 值由原始 SVM 模型的 0.48 显著提升至 0.70，提升幅度达到 45.8%。此外，为进一步分析各变量对预测模型的贡献，比较了 aspatial RF 与 SDO-RF

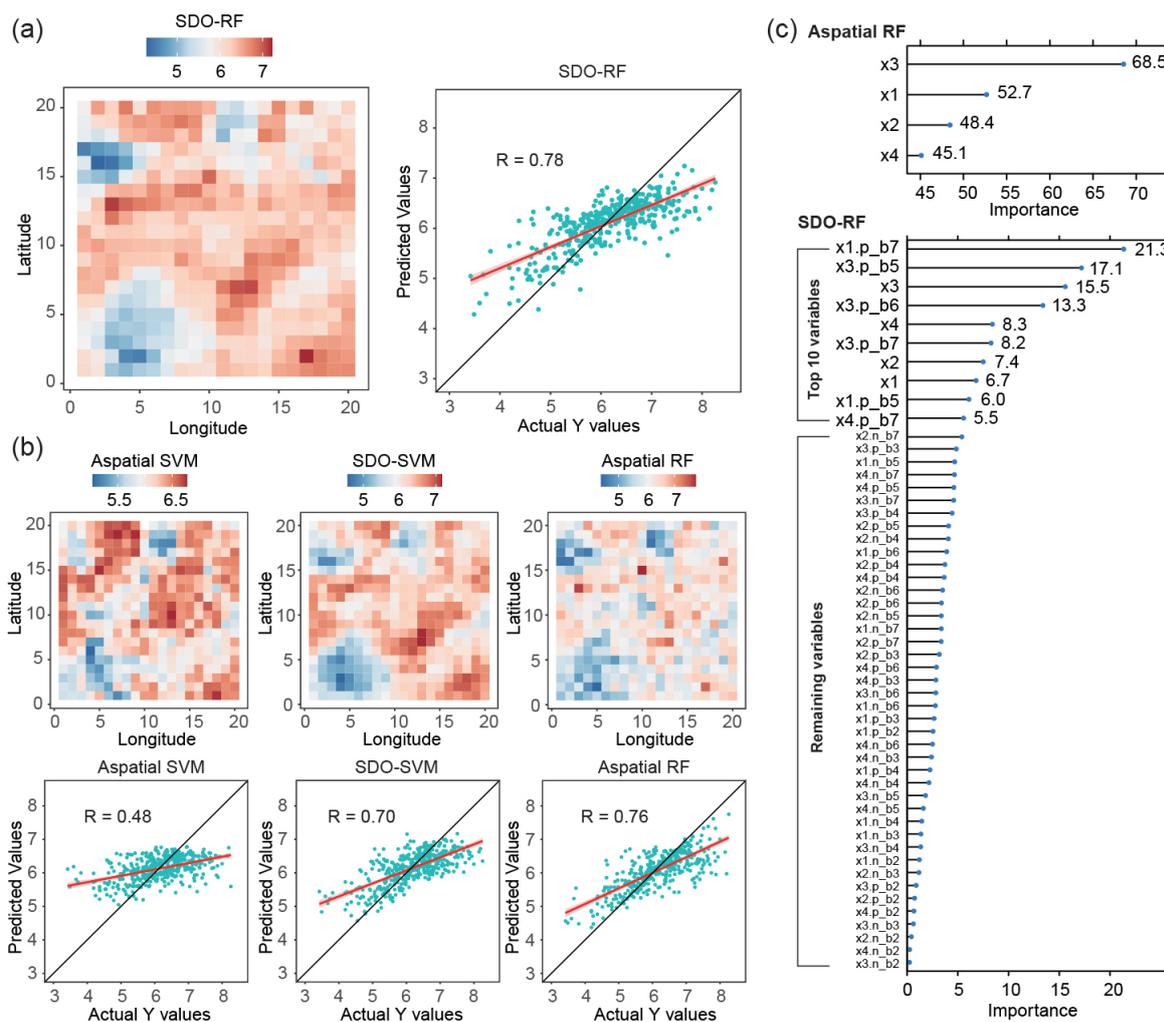


图 4-7 SDO 模型在模拟数据上的预测结果，采用随机森林 (RF) 和支持向量机 (SVM) 进行分析。(a) 基于 RF 的 SDO 预测结果；(b) 基于 RF 和 SVM 的非空间预测结果，以及 SDO-SVM 预测结果；(c) 变量重要性分析，其中上方为非空间 RF 模型，下方为 SDO-RF 模型。

Fig. 4-7 Results of SDO modeling on simulation data using random forest (RF) and support vector machine model (SVM). (a) SDO prediction result using RF, (b) Aspatial prediction results based RF and SVM, and SDO-SVM prediction, (c) Importance of the individual variables for the aspatial (upper) and SDO (lower) RF models.

模型中各变量的重要性（见图 4-7 (c)）。结果显示，在两个模型中， x_3 均为最具影响力的变量；特别是在 SDO-RF 模型中，前十大变量中有四个为 x_3 及其派生的 SDO 变量，进一步强调了空间异常值信息对提升模型预测能力的重要作用。

上述模拟实验结果表明，SDO 模型能够基于局部空间异常信息进行精确回归预测。在预测精度、结果平滑性、异常识别能力及模型拟合度方面，SDO 模型整体优于传统非空间预测方法。

4.4.2 小麦产量的变量选择与相关性分析

本章开发的 SDO 模型支持不同空间尺度下的数据建模与训练。数据预处理包括缺失值处理、生成 1000 米分辨率的预测网格，并对每个网格单元进行解释变量的重采样处理。

首先对全部变量进行了相关性分析，结果如图 4-8 所示。分析结果表明，部分变量与响应变量（小麦产量）之间具有显著相关性，特别是气温、蒸散发、EVI、NDVI、总降水量、粉粒（Silt）与砂粒（Sand）。初步阶段曾将风速（WS）作为候选解释变量，但相关性分析显示风速与小麦产量及其他变量之间的相关性极低，故在最终建模中予以剔除。

根据 SDO 模型框架及样本点与预测网格的位置分布，识别不同缓冲尺度下解释变量的空间异常值，并计算对应的异常强度指数（Outlier Strength Index, OSI），进而生成 SDO 变量。生成的 SDO 变量分为正向与负向变量：正向变量代表模型识别出的正向异常值，负向变量代表负向异常值。

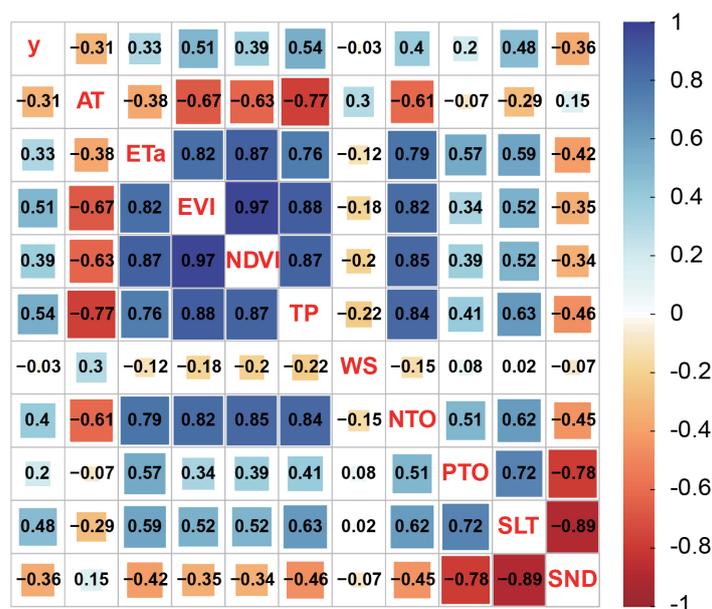


图 4-8 小麦产量与解释变量的相关性分析，包括气温、蒸散量、EVI、NDVI、总降水量、总氮、总磷、粉粒和砂粒。

Fig. 4-8 Correlation analysis for wheat production and explanatory variables, including air temperature, evapotranspiration, EVI, NDVI, total precipitation, total nitrogen, total phosphorus, silt, and sand.

4.4.3 生成 SDO 变量和基于 SDO 模型构建机器学习预测模型

根据 SDO 模型的概念和原来，生成了所有解释变量的第二维度异常变量，称为 SDO 变量。图 4-9 展示了气温变量在不同缓冲尺度下生成的 SDO 变量。从图中可以看

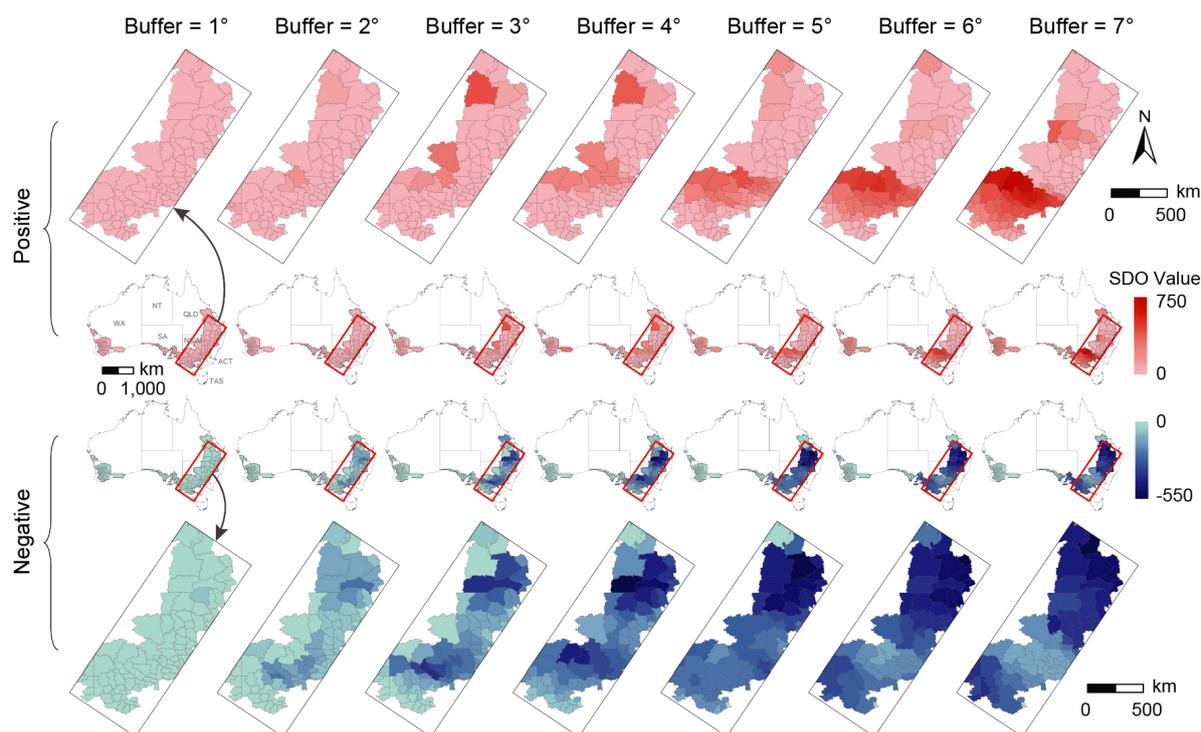


图 4-9 不同缓冲区下 SDO 模型生成的 SDO 温度变量分布，其中红色表示正异常点，蓝色表示负异常点。

Fig. 4-9 The distribution of the second-dimension temperature variable generated by the SDO model under different buffers, which includes positive outliers (red) and negative outliers (blue).

出，气温的异常值主要分布于澳大利亚东部区域，尤其集中在新南威尔士州（NSW）、维多利亚州（VIC）与昆士兰州（QLD）部分地区。这种空间特征与该区域复杂的山地地形、东南信风与洋流共同作用密切相关。

图 4-10 展示了总降水量、蒸散发、NDVI、EVI、总氮、总磷、砂粒与粉粒等多个解释变量的 SDO 变量生成结果。与图 4-9 类似，该图呈现了各缓冲尺度下识别的正向与负向异常值。随着缓冲半径的扩大，异常值的强度也相应增加，表明局部空间模式的偏离程度在较大范围下更为显著。接着将所有 SDO 变量与原始解释变量联合输入支持向量机（SVM）进行建模训练，初步结果验证了 SDO 模型在提升空间异常识别能力与整体预测精度方面的有效性。

4.4.4 SDO 模型验证与敏感性分析

SDO 模型进一步与七种主流机器学习方法集成，构建出一组 SDO 增强型模型。模型性能采用五折交叉验证（five-fold cross-validation）进行评估，每轮使用 80% 的数据用于训练，20% 用于验证，最终取五轮结果的平均值作为精度指标。表 4-3 对比了各模型在决定系数 R^2 、均方根误差（ $RMSE$ ）与平均绝对误差（ MAE ）方面的表现。结果表明，引入 SDO 模型后，所有模型的 R^2 均明显提高， $RMSE$ 与 MAE 均显著下

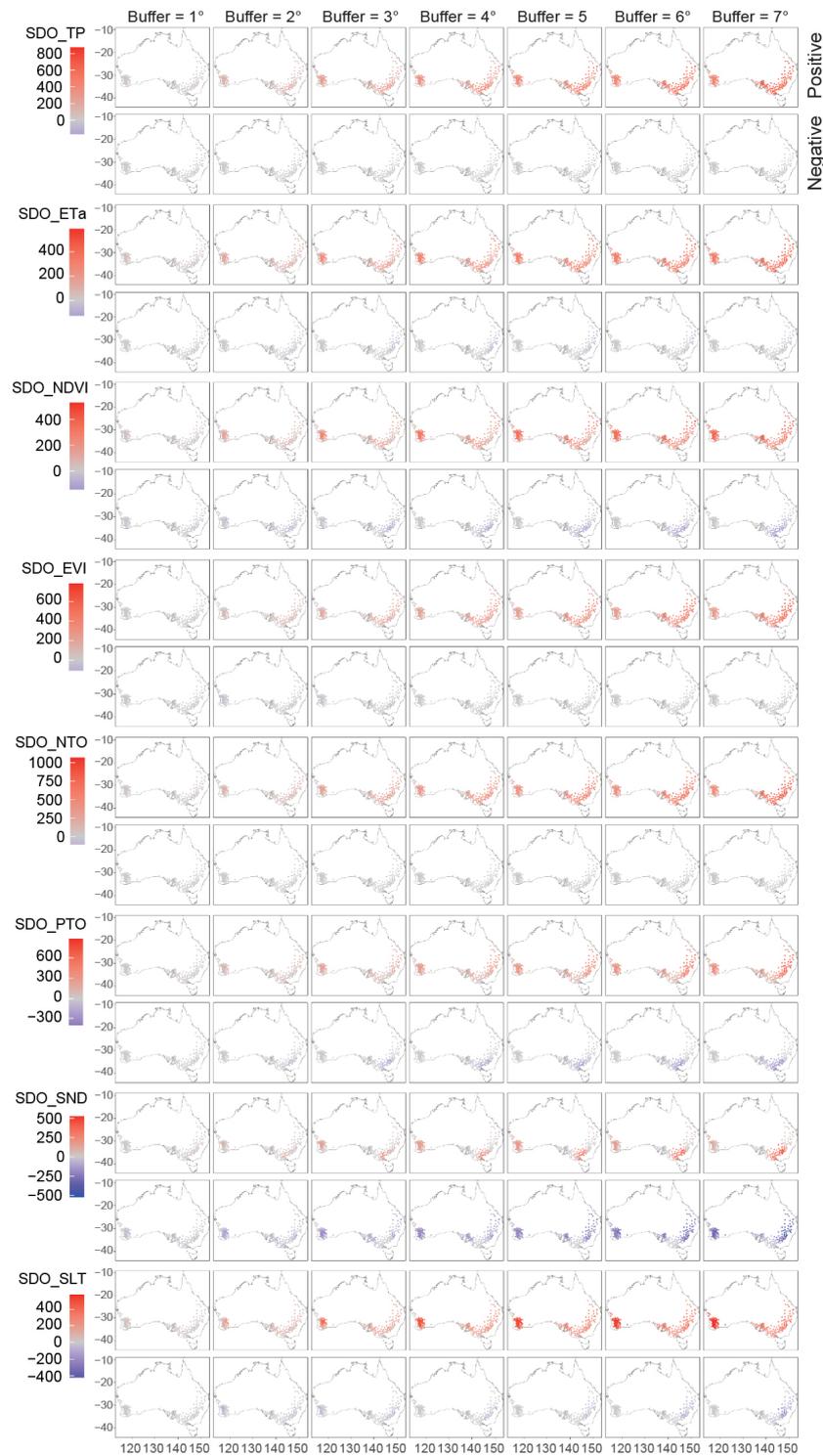


图 4-10 SDO 模型生成的解释变量的第二维度异常点分布，包括总降水量、蒸散量、NDVI、EVI、总氮、总磷、砂粒和粉粒。

Fig. 4-10 Distributions of the second-dimension outlier of explanatory variables, including total precipitation, evapotranspiration, NDVI, EVI, total nitrogen, total phosphorus, sand, and silt generated by SDO model.

降。其中，SDO-SVM 模型综合性能最佳， R^2 达到 0.713，优于 SDO-KNN (0.643) 与 SDO-RF (0.664)；在 $RMSE$ 方面，SDO-SVM 误差最小 (0.291)，其次为 SDO-RF (0.321) 与 SDO-KNN (0.322)；而在 MAE 方面，SDO-SVM (0.218)、SDO-KNN (0.217) 与 SDO-XGB (0.239) 表现最优。总体而言，SDO 增强模型在所有指标上均优于其非空间对应模型，其中 SDO-SVM 展现出最佳综合性能，验证了 SDO 模型在捕捉传统模型忽略的空间异常信息方面的优势。

表 4-3 SDO 模型相较于非空间模型在机器学习中提升模型精度的对比分析（加粗值表示表现最优的 SDO 结果）。

Tab. 4-3 Improvements of model accuracy in machine learning by SDO models compared with aspatial models (bold values indicating the best-performing SDO results).

| | Model | RF | CRM | XGB | SVM | GBM | KNN | ENR |
|--------|-------------|-------|--------------|-------|--------------|-------|-------|-------|
| R^2 | Aspatial | 0.500 | 0.653 | 0.504 | 0.532 | 0.549 | 0.479 | 0.558 |
| | SDO | 0.664 | 0.684 | 0.643 | 0.713 | 0.657 | 0.643 | 0.614 |
| | Improvement | 32.6% | 4.8% | 27.6% | 33.9% | 19.8% | 34.4% | 10.1% |
| $RMSE$ | Aspatial | 0.383 | 0.321 | 0.381 | 0.368 | 0.369 | 0.386 | 0.360 |
| | SDO | 0.321 | 0.308 | 0.321 | 0.291 | 0.316 | 0.322 | 0.335 |
| | Reduction | 16.3% | 4.2% | 15.9% | 20.9% | 14.5% | 16.7% | 7.1% |
| MAE | Aspatial | 0.288 | 0.243 | 0.286 | 0.271 | 0.284 | 0.299 | 0.292 |
| | SDO | 0.247 | 0.217 | 0.239 | 0.218 | 0.239 | 0.244 | 0.249 |
| | Reduction | 14.3% | 10.6% | 16.6% | 19.6% | 15.8% | 18.3% | 14.6% |

为评估 SDO 模型的稳健性，本研究通过调整缓冲间隔、缓冲阈值以及异常值阈值（标准差，SD）开展敏感性分析。图 4-11 给出了不同参数设置下的交叉验证均方根误差（CV-RMSE）。在图 4-11(a) 中，缓冲间隔固定为 1° ，缓冲阈值分别设为 6° 、 7° 和 8° 。结果显示 CV-RMSE 变化幅度极小，其中阈值为 7° 时误差最低 (0.291)，较 6° 降低 1.17%，较 8° 增加 0.83%。图 4-11(b) 将缓冲阈值固定为 7° ，对缓冲间隔进行测试：当间隔为 2° 时，CV-RMSE 较 1° 提高 2.21%；当间隔为 3° 时，则较 1° 下降 1.31%。图 4-11(c) 评估了模型对异常值阈值选择的敏感性，分别测试 1.5 SD、2 SD 和 2.5 SD，CV-RMSE 介于 0.291–0.299 之间，其中 2 SD 时误差最低。图 4-11(d) 进一步展示了各参数设置下 CV-RMSE 的百分比变化，其波动均控制在 $\pm 5\%$ 以内。综合上述结果，表明 SDO 模型具有良好的稳健性，其预测性能在不同空间参数配置下保持稳定；同时，第二维度异常变量的引入有效提升了模型的泛化能力，而不会过度依赖特定的缓冲参数选择。

4.4.5 小麦产量的空间预测与评估

图 4-12 展示了非空间 SVM 模型 (aspatial SVM) 与第二维度异常支持向量机 (SDO-SVM) 模型的空间预测结果对比。与传统非空间模型相比，SDO-SVM 所生成的预测结

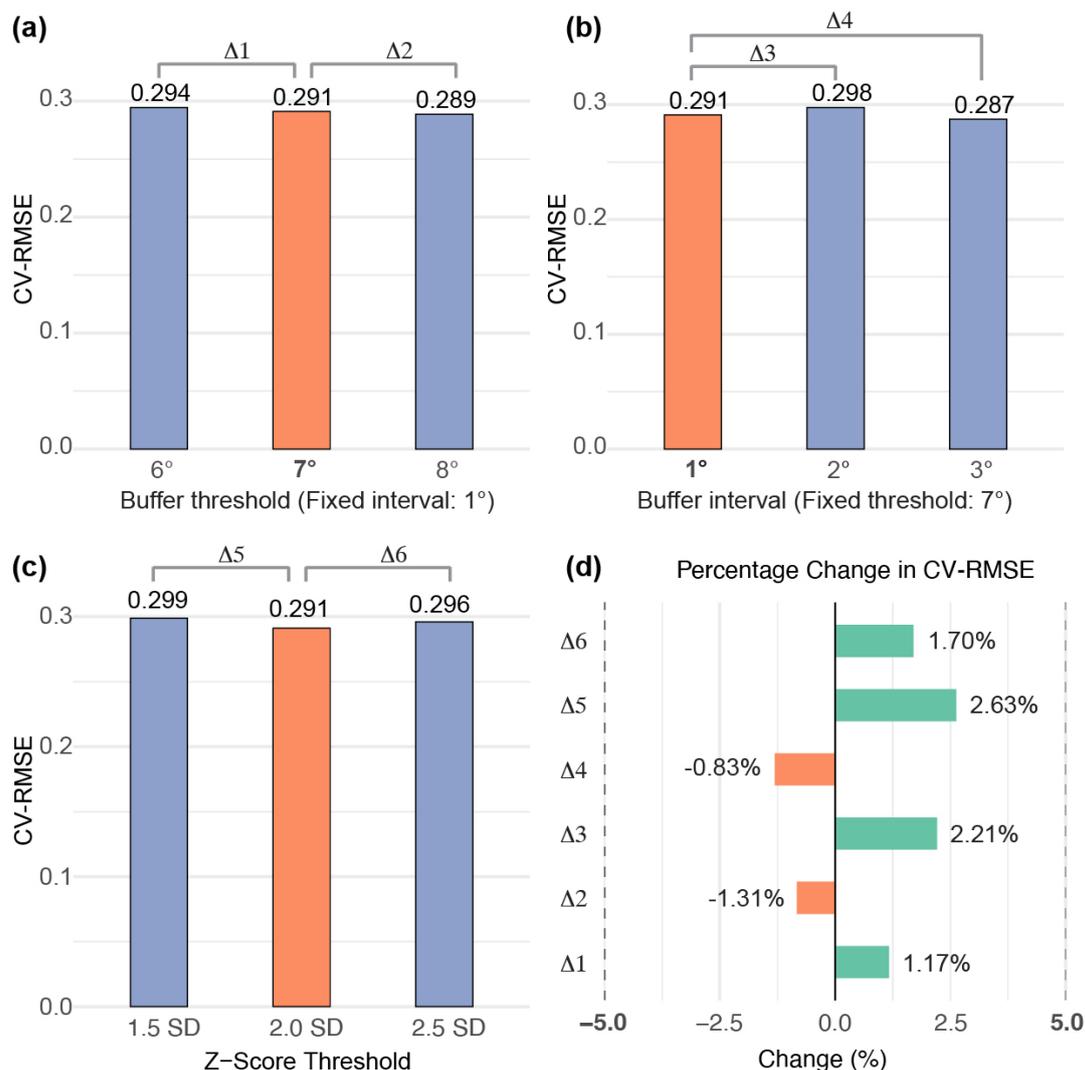


图 4-11 SDO 模型敏感性分析：缓冲阈值、缓冲间隔与异常值（标准差）阈值对模型精度（交叉验证均方根误差，CV-RMSE）的影响

Fig. 4-11 Sensitivity analysis of the SDO model: impact of buffer threshold, buffer interval, and outlier (standard deviation) threshold settings on model accuracy (cross-validation root mean square error, CV-RMSE).

果在空间连续性方面更为优越，能够在保持广域空间结构特征的同时，更有效地融合局部异常信息（图 4-12 (c)）。相较之下，*aspatial SVM* 模型的预测图则表现出较为粗糙和不均匀的空间变异特征（图 4-12 (a)）。通过 R^2 、 $RMSE$ 和 MAE 三项指标对两种模型进行定量比较，进一步验证了 SDO 模型的性能优势。具体而言，SDO 模型的 R^2 从 0.531 提升至 0.713，增幅达 33.9%；同时， $RMSE$ 和 MAE 分别下降了 20.9% 和 19.6%。

值得强调的是，这两种模型使用的均为同一组解释变量，且均来源于相同的环境数据产品。两者的关键差异在于 SDO 模型引入了额外的“第二维度”变量，用于量化预测变量在未采样邻域中的正向和负向异常。这些异常变量有效捕捉了局地空间模式，使

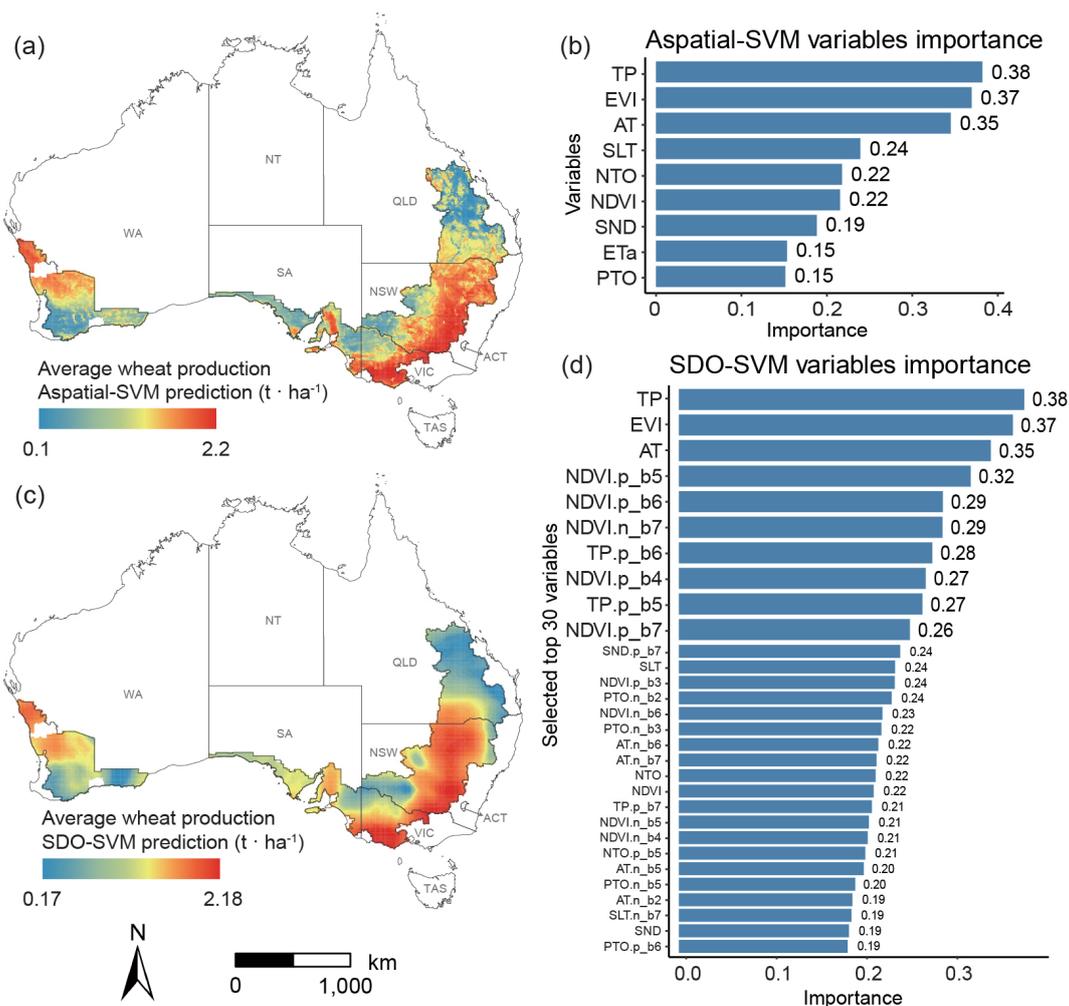


图 4-12 澳大利亚小麦平均产量预测：基于非空间模型 (a) 和 SDO 机器学习模型 (c) 的结果。机器学习模型中前 30 个 SDO 变量 (d) 和非空间变量 (b) 的重要性分析。

Fig. 4-12 Prediction of average wheat production in Australia based aspatial model (a) and SDO machine learning model (c). The importance of top 30 SDO variables (d) and aspatial variables (b) in the machine learning model.

模型更好地表征空间非平稳性与极端事件的影响。因此，模型性能的提升并非源于平均效应的增强，而是得益于异常变量所引入的空间结构信息。

虽然两种模型在变量构建中的主导变量基本一致，但其变量重要性表现仍存在一定差异。图 4-12 (b) 展示了非空间 SVM 模型的变量重要性，图 4-12 (d) 则展示了 SDO-SVM 模型中贡献最高的前 30 个 SDO 变量。在两个模型中，降水、EVI 与气温为前三位的重要变量，重要性分别为 0.38、0.37 与 0.35。然而，非空间模型未考虑来自样本点之外的空间信息，可能忽视了局部异常的影响，从而限制预测精度。

相较而言，SDO 模型通过引入不同空间尺度下的第二维度异常变量，增强了与观测点之间的空间关联性。这些变量不仅影响周边样本点的值，更对识别空间异常具有关键意义。如图 4-12 (d) 所示，除降水、EVI 与气温外，不同缓冲尺度下的 NDVI 与

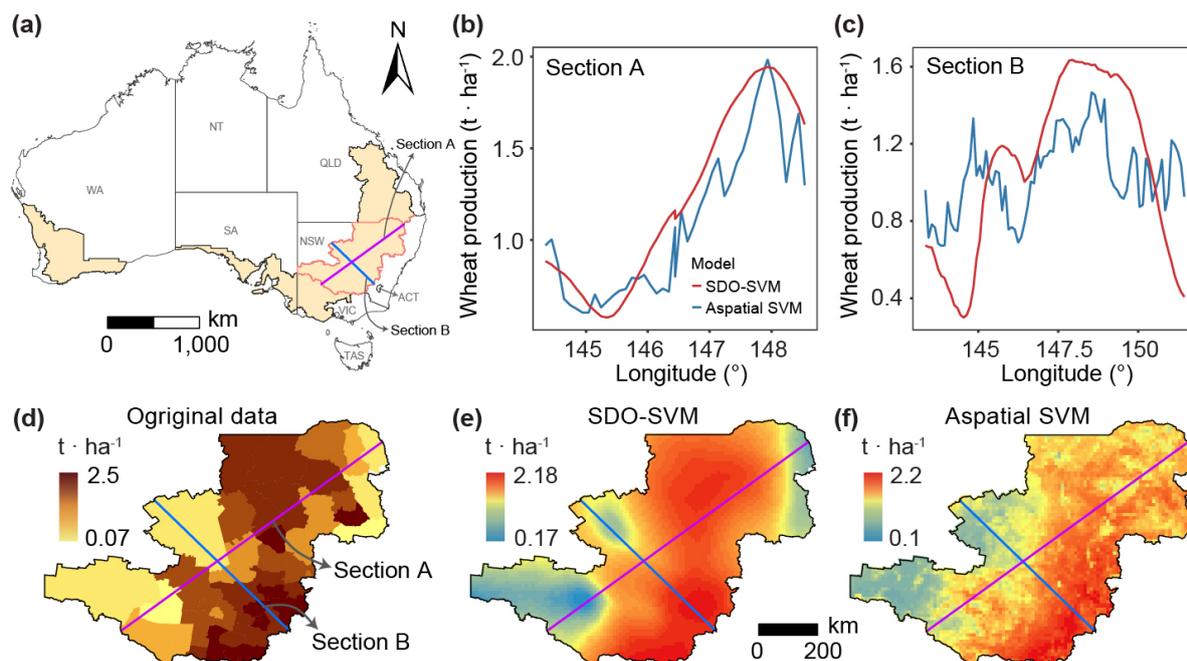


图 4-13 新南威尔士州小麦带区域 SDO 模型与非空间模型预测对比分析。(a) 新南威尔士州小麦带区域内截面 A (西南至东北) 和截面 B (西北至东南) 的位置。(b, c) 沿截面 A 和 B 的小麦产量预测, 对比 SDO-SVM 与非空间 SVM 模型。(d) 原始小麦产量数据, 并叠加载面。(e, f) SDO-SVM 模型 (e) 和非空间 SVM 模型 (f) 预测的小麦产量, 叠加载面。

Fig. 4-13 Comparison of SDO and aspatial model predictions along cross-sections in the wheat belt region of New South Wales. (a) Locations of section A (southwest to northeast) and section B (northwest to southeast) in the wheat belt region of New South Wales. (b, c) Predicted wheat production along sections A and B, comparing SDO-SVM and aspatial SVM models. (d) Original wheat production data with cross-sections overlaid. (e, f) Predicted wheat production from the SDO-SVM model (e) and the aspatial SVM model (f), with overlaid cross-sections.

降水的 SDO 变量也表现出较高的重要性。例如, NDVI 在 500km 与 600km 缓冲尺度下的正向异常变量重要性分别为 0.32 与 0.29, 而 700km 缓冲尺度下的负向异常变量为 0.29; 类似地, 600km 缓冲下的降水 SDO 变量贡献值达 0.28。结果表明, SDO 模型生成的异常变量在模型构建与性能提升中起到关键作用, 显著增强了模型的预测准确性与鲁棒性。

为进一步评估 SDO 模型对空间异常的识别能力, 本文选取新南威尔士州 (NSW) 小麦带的两个剖面进行对比分析 (图 4-13)。剖面 A (自西南向东北) 与剖面 B (自西北向东南) 代表小麦产量空间变化的典型方向。图 4-13 (a) 展示了其在澳大利亚及小麦带内的位置, 图 4-13 (b) 与 (c) 分别展示了 SDO-SVM 与非空间 SVM 模型在两个剖面上的预测趋势。尽管两者总体趋势相似, SDO-SVM 更准确地捕捉到极值波动, 尤其是在 145°E 至 146°E 之间的急剧下降与 147.5°E 至 148°E 之间的平稳上升; 非空间模型则表现出更多突变与噪声。图 4-13 (d)-(f) 展示了剖面预测与原始数据的对比, SDO 模型在增强空间连续性的同时保持了预测精度。

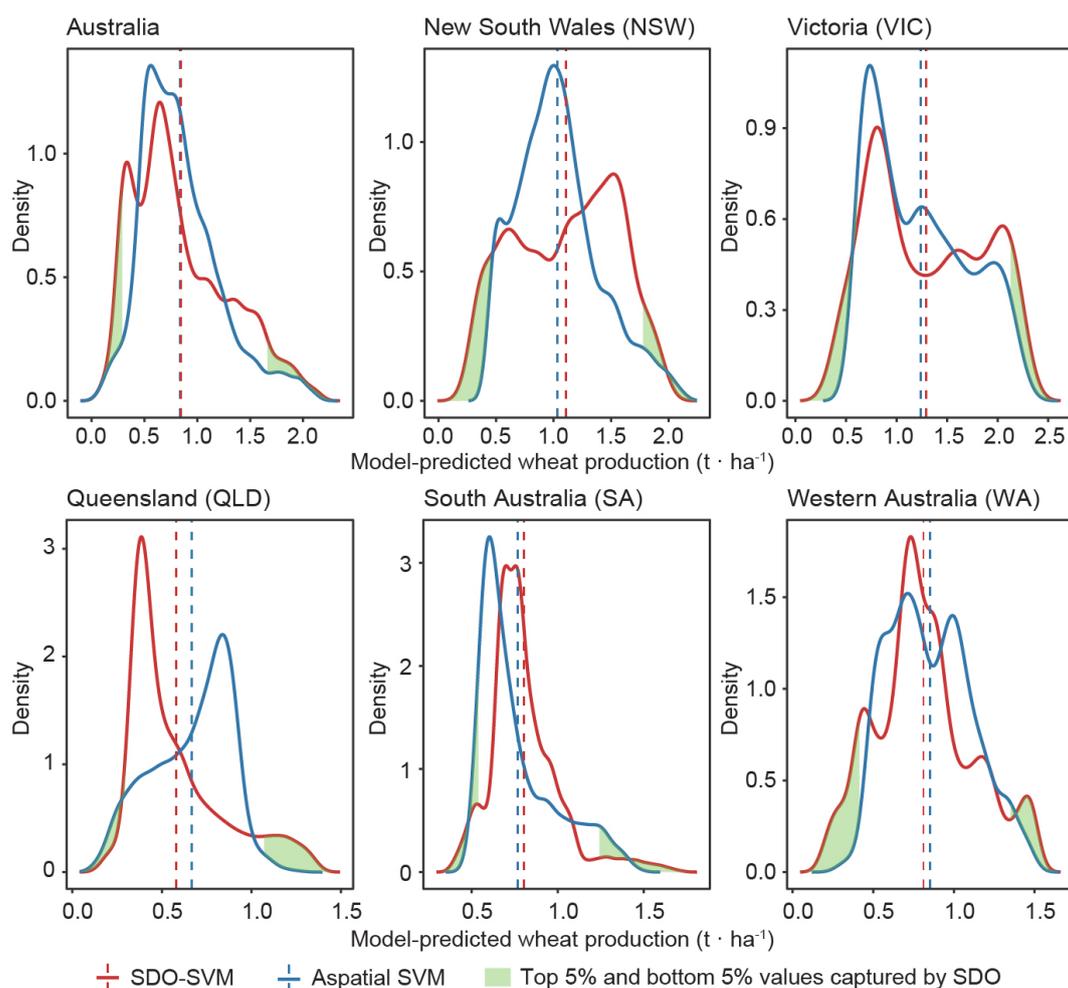


图 4-14 SDO 模型与非空间模型在澳大利亚及各州的平均小麦产量预测结果的密度分布图。

Fig. 4-14 Density distribution map of the average wheat production prediction results of the SDO and aspatial models in Australia and each state.

图 4-14 比较了 SDO 与非空间模型在澳大利亚整体及各州范围内预测的小麦产量密度分布。整体而言，两种模型在全澳小麦带范围的密度分布较为一致，均值与尾部分布相近，表明模型预测在全区范围具备一致性。然而，在分州分析中，SDO 模型（红色曲线）在分布两端密度更高，表明其能更有效检测出正向与负向异常值。特别是在 NSW、VIC 与西澳州（WA）中，负向异常值的识别更为明显，而在 QLD 中则更多识别出正向异常值，进一步体现了 SDO 模型在捕捉极值方面的敏感性与优势。

4.5 讨论

本研究引入了第二维度异常（Second-Dimension Outliers, SDO）的概念，并基于 SDO 构建了一系列机器学习空间预测模型。通过纳入样本点之外的异常信息，这些模型能够提高预测精度并减少模型误差。SDO 模型在因变量与预测因子在样本外位置的

异常特征之间建立联系，从而实现对空间异常的更全面表征。通过在多个缓冲距离上生成局地 SDO 变量，模型捕捉到常规非空间模型往往忽视的复杂地理与环境格局。

空间异常在提升预测精度方面发挥着关键作用，可通过捕捉异常空间模式、解释局地极值、减轻空间异质性带来的误差并增强模型稳健性 (Baba et al. 2022; Lu et al. 2009)。空间自相关表明邻近区域的观测值趋于相似 (Overmars et al. 2003; Song 2023)，而空间异常则体现对这一规律的偏离，使模型能够基于局地异常而非统一的空间依赖进行修正 (Lu et al. 2009)。这对于捕捉诸如气候异常导致的农业产量波动等极端事件尤为重要 (Anselin 2019)。此外，空间异常反映了解释变量（如土壤性质、植被指数与气候因子）在不同区域的差异，使模型得以更准确地校正并优化预测 (Cai et al. 2024)。通过引入这一额外的空间信息层，SDO 模型显著提升了预测精度并降低了不确定性 (Shen et al. 2021)。

SDO 方法的关键技术特点在于其对缓冲区选择与异常值定义的依赖，这些参数直接影响模型性能。为评估其敏感性，本研究开展了综合敏感性分析，考察不同缓冲阈值 m 及缓冲间隔对模型精度的影响。结果表明，以交叉验证 RMSE 衡量的模型性能在不同参数配置下波动均小于 5%，验证了 SDO 方法在多种缓冲设置下的稳健性。缓冲区大小需在空间分辨率与计算效率之间取得平衡：较大的缓冲可捕捉更广泛的空间结构，但可能稀释局地模式；较小的缓冲强调局地差异，但也可能忽略更广的背景影响。经验结果表明，缓冲阈值 7° （约 700 km）与 1° 间隔（约 100 km/步）在多尺度空间表达与计算可行性之间达成了合理折衷。同样，异常值阈值的选择对 SDO 变量构建具有重要影响。本研究采用 $\bar{x} \pm 2\sigma$ 阈值识别异常值，以系统且客观的方式检测异常，同时避免对微小波动过度敏感。

虽然本研究选择 5-10 个缓冲区、覆盖最大样本间距 10-20% 的范围是基于经验惯例，并得到了近期研究的支持（如城市与环境研究中常用的 100-500 m 或 250-1,000 m 缓冲；(Qi et al. 2022; Shi et al. 2021; Zhang et al. 2021)），但该范围仍具有启发性。最优缓冲大小实际上取决于具体的空间过程——例如扩散距离或政策驱动尺度——可通过自动校准（如 GWR/MGWR 的交叉验证）或自适应、位置特定的缓冲进一步优化。因此，建议未来工作探索缓冲大小优化技术与可变缓冲方案，以更好地将 SDO 方法适配于多样化的数据集和应用场景。

第二维度异常 (SDO) 概念通过引入多尺度空间特征，扩展了传统空间异常值分析的能力，增强了模型同时捕捉局地变异与更大尺度空间关联的能力。SDO 通过多缓冲区尺度提取空间异常，克服了单尺度方法的局限性，有效地提升了空间异常的表征精度。与传统的基于邻域平均或加权的模型不同，SDO 选择性地提取在缓冲区内更可能对因变量产生显著影响的异常值，从而保留了局地极值信号，避免了整体邻域平均

可能带来的异常信号弱化。尽管完整聚合空间上下文在某些情况下可能提升预测性能，但也可能引入冗余信息、掩盖局地偏离，甚至增加模型复杂度。研究表明，SDO 通过在不同距离尺度下独立提取异常值，能够在不增加不必要复杂性的前提下，有效而准确地刻画空间异质性。未来的研究可考虑结合异常值驱动与全邻域策略的混合框架，以进一步提升模型在多样化空间场景下的预测表现。

将 SDO 模型应用于澳大利亚小麦产量预测表明，多个解释变量在空间上存在强烈的局地异常，这些异常对产量变异具有显著影响。传统非空间模型难以识别这些异常，导致预测精度降低。而在引入多尺度空间异常后，SDO 模型在预测性能方面表现出明显提升。例如，在 500 km 缓冲区范围内，NDVI 的正异常贡献占比达到 32%，而其他尺度的异常信号则更为显著。SDO 模型能够有效识别并利用这些空间异常，从而提供更具情境感的空间变异表征。上述发现进一步验证了 SDO 在捕捉与利用空间异常方面的有效性，有助于提升空间预测的稳健性、精度与可解释性。

SDO 的引入为空间建模提供了一种有效机制，能够通过利用多尺度空间变异来提升预测精度。研究表明，空间异常在优化模型性能方面具有关键作用，特别是在应对极端值、空间异质性和非线性空间依赖方面。此外，灵敏度分析结果也表明，SDO 对不同缓冲参数与异常阈值设置具有良好的稳定性，进一步体现了其鲁棒性与适应性。总体而言，SDO 框架对现有空间预测方法构成了重要补充，提供了一种更全面、更具可扩展性的思路，以应对空间异常与不确定性，并提升模型在多地理应用场景中的可靠性。

4.6 小结

本章提出了第二维度异常 (Second-Dimension Outliers, SDO) 的概念，基于样本点的空间位置，在多尺度局地缓冲区内识别正向与负向异常，并进而构造 SDO 变量，发展了第二维度异常机器学习模型用于精确预测局地异常。通过前述空间模拟数据和澳大利亚小麦产量预测的案例分析，验证了该模型在多尺度捕捉空间异常方面的优越性能。结果表明，引入样本点之外的地理与环境特征显著提升了预测精度，避免了异常信息的模糊化，从而有效降低了模型误差。

第五章 基于空间异常的异质性量化解释模型

本章基于上一章提出的第二维度异常模型 (SDO)，提出了第二维度异常驱动的异质性量化模型 (SOH)。该模型旨在通过量化空间模式变量的异常信息，来解释目标对象的空间分布模式，分析空间现象的异质性。SOH 模型不仅能够识别空间数据中的异常模式，还能为空间决策提供有力支持。同时，将这两个模型应用于澳大利亚小麦和大麦产量数据，进一步验证了 SOH 在空间数据异质性量化和解释方面的能力。

5.1 概述

地理过程往往呈现显著的空间非平稳性与多尺度异质性；然而主流的空间预测与解释框架（无论是传统统计模型、地理加权方法，还是机器学习/深度学习驱动的空间建模）通常依赖采样点处的原始环境因子，或其经空间聚合后的均值指标，在模型训练中将局部极值、突变边界与稀有但高影响力的异常格局视作噪声加以平滑或剔除。这种处理方式容易掩盖“小区域强信号”对宏观空间分布的驱动作用，导致对空间分异机制与主导因子的识别不充分，进而限制模型在区域管理与精细化农业决策中的应用潜力。基于上一章提出的第二维度异常 (SDO) 思想，本章进一步提出：能否将多尺度邻域内的空间异常信息系统性量化，并转化为可与原始解释变量协同建模的“空间模式变量”，以提升空间异质性解释力。围绕这一核心问题，本章发展了第二维度异常驱动的异质性量化模型 (Second-Dimension Outlier-driven Heterogeneity, SOH)。

该模型旨在从空间模式视角定量解释地理现象的空间分布异质性，通过集成多尺度缓冲区内的空间异常信息，将传统的空间解释变量扩展为包含空间局部极值与结构突变的“空间模式变量” (Spatial Outlier Patterns, SOPs)，并结合空间判别框架评估其对目标变量空间分布的解释力。SOH 模型不仅能识别个体变量的空间主导程度，还能系统刻画变量之间的非线性交互作用及其在不同类别和不同区域的空间表达特征，提升了空间异质性分析的深度。

最后，分别通过澳大利亚小麦产量、大麦产量为案例进行实证分析，本章验证了 SOH 模型在识别主导因素、量化空间分异机制及支撑空间决策方面的有效性。该方法为理解复杂地理过程提供了一种新的建模路径，也为面向空间认知和区域管理的实践提供了理论支持与技术手段。

5.2 材料与方法

5.2.1 数据及预处理

本章的研究区域和空间数据与上一章一致，仍然选取澳大利亚小麦主产区（见图 4-2）。研究单元依旧选择地方政府区域（Local Government Areas, LGAs）区域，2021 年有 179 个 LGA 区域有官方公布的小麦产量数据，因此选择 179 个该区域的平均小麦产量数据作为研究对象。

空间解释变量数据包括气候变量、土壤属性、环境变量三大类，包含 9 个变量，具体数据来源和预处理方法同前一章。

计算空间模式变量（SOPs）时采用的网格数据集同样使用 1000 米分辨率，既保证了空间分辨率的适当性，又能有效捕捉到小麦产量的空间分布特征。LGA 区域内的网格数据使用该空间区域的统一平均值，而待分析格网上的数据则是重采样到 1000 米分辨率上的格网上的，每个格网对应的值不同。

5.2.2 SOH 模型：异常值驱动的异质性量化模型

5.2.2.1 SOH 模型

SOH 模型的构建包括结合原始解释变量与第二维度异常模式（SOPs），并集成决策树和地理探测器来共同衡量空间模式变量对目标变量的解释力。其中 SOPs 是通过第二维度异常（SDO）模型生成的，其包含有目标变量一定邻域范围内的正负异常值信息，所包含的空间信息远远大于原始变量，因而 SOPs 空间模式有着更强的解释力。SOH 模型的具体操作步骤如图 5-1：

首先根据经验和文献资料收集了影响小麦产量的解释变量数据，这些数据经过重采样、裁剪、极端值处理和标准化等预处理，即可作为原始的变量（Vars）。这些变量在以 0.1° 为缓冲区间隔，从 0.1° 到 1° 的缓冲区内使用 SDO 模型进行计算，得到一系列的空间模式变量（SOPs）。这些 SOPs 包含了目标变量在不同缓冲区内正负异常值信息。接着，将原始变量与 SOPs 结合，集成决策树和地理探测器，构建了 SOH 模型来评估这些变量对目标变量的解释力。通过不同的变量组合与分类，分别分析了不同情况下的 SOPs 和 Vars 对空间异质性的解释力度。

5.2.2.2 第二维度异常驱动的异质性（SOH）模型概念

第二维度异常驱动的异质性（SOH）模型旨在通过量化空间模式变量的异常信息，来解释目标对象的空间分布模式，分析空间现象的异质性。SOH 模型基于第二维度异常（SDO）模型，通过集成多尺度缓冲区内空间异常信息，将传统的空间解释变量扩展为包含空间局部极值与结构突变的“空间模式变量”（SOPs），并结合决策树和地理探测器模型评估其对目标变量空间分布的解释力。其主要原理和公式如下：

设原始空间变量定义如下：

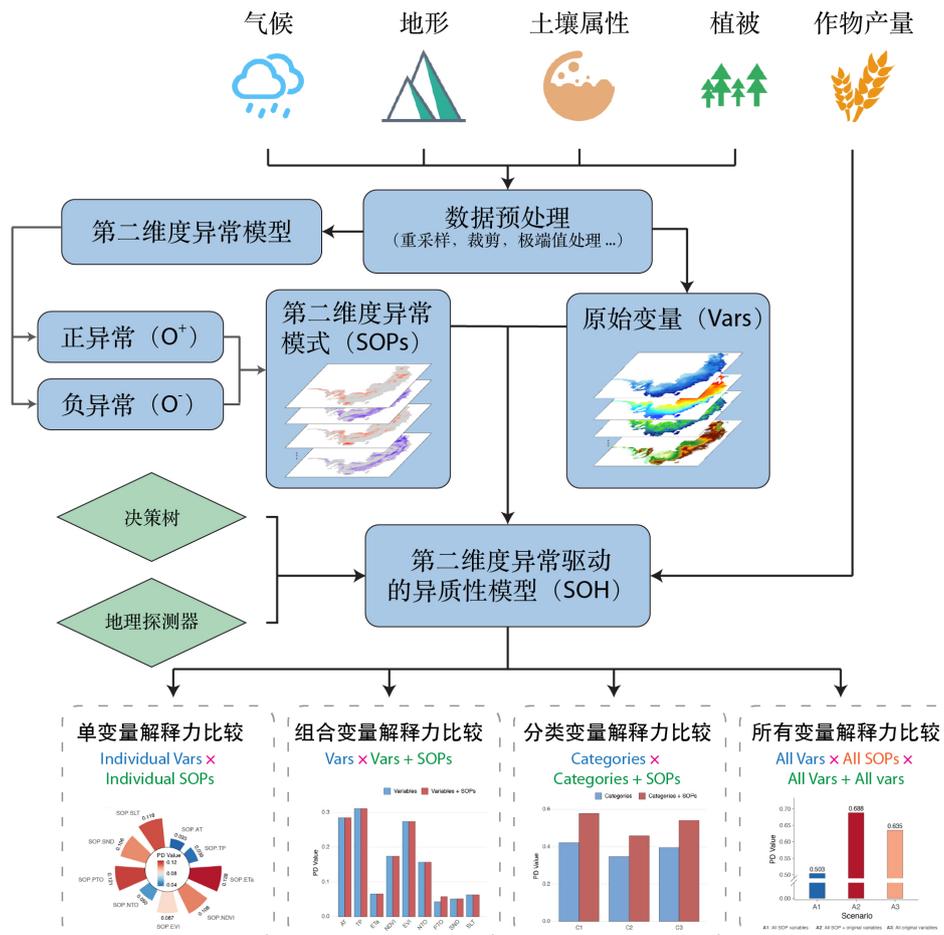


图 5-1 第二维度异常驱动的异质量化 (SOH) 模型框架与流程示意图
 Fig. 5-1 Framework and Workflow of the Second-Dimension Outlier-driven Heterogeneity (SOH) Model

$$X = \{x_i\}_{i=1}^n \quad (5-1)$$

式中, x_i 表示变量 X 在空间单元 i 处的取值, n 为空间单元的总数。

将空间模式变量 Ψ 定义为一组由第二维离群信息构建而成的变量集合:

$$\Psi = \bigcup_{r \in R} \{O_+^{(r)}, O_-^{(r)}\} \quad (5-2)$$

式中, $R = \{r_1, r_2, \dots, r_k\}$ 表示一组缓冲区半径, $O_+^{(r)}$ 和 $O_-^{(r)}$ 分别代表在缓冲距离 r 下从原始变量 X 中识别出的正向和负向第二维空间离群变量。

对每个空间单元 i , 其对应的空间模式向量 Ψ_i 可表示为:

$$\Psi_i = [O_{+,i}^{(r_1)}, O_{-,i}^{(r_1)}, O_{+,i}^{(r_2)}, O_{-,i}^{(r_2)}, \dots, O_{+,i}^{(r_k)}, O_{-,i}^{(r_k)}] \quad (5-3)$$

式中, $O_{+,i}^{(r)}$ 和 $O_{-,i}^{(r)}$ 分别表示在缓冲距离 r 下, 空间位置 i 周围的正向和负向离群强度或存在性。

基于地理探测器模型, 空间模式变量 Ψ 的解释力 (Power of Determinant, PD) 定义如下:

$$PD = \Omega_{\Psi} = 1 - \frac{\sum_{h=1}^L N_h^{(\Psi)} \cdot \sigma_h^{2(\Psi)}}{N^{(\Psi)} \cdot \sigma^{2(\Psi)}} \quad (5-4)$$

式中, Ω_{Ψ} 表示因子对空间模式变量 Ψ 总体方差的解释程度 (取值范围为 0 到 1), L 为将解释变量划分的类别 (或层) 数, $N_h^{(\Psi)}$ 为第 h 层中的空间单元数量, $N^{(\Psi)} = \sum_{h=1}^L N_h^{(\Psi)}$ 为空间单元总数, $\sigma_h^{2(\Psi)}$ 表示第 h 层中变量 Ψ 的层内方差, $\sigma^{2(\Psi)}$ 为研究区域中变量 Ψ 的总体方差。

通过解释力 PD 的计算, SOH 模型能够量化空间模式变量对目标变量空间分布模式的解释能力, 从而揭示空间现象的异质性。

5.2.2.3 模型验证

为了验证 SOH 模型在空间异质性量化与解释方面的能力和有效性, 以澳大利亚小麦产量作为案例进行实证分析。通过对比不同变量组合下的解释力 (PD) 值, 评估 SOH 模型在识别主导因素、量化空间分异机制及支撑空间决策方面的表现。

主要的模型验证形式如下:

- (1) 单个变量解释力分析 (Individual SOP vs Individual Var)
 - (a) 单个原始变量的解释力;
 - (b) 单个 SOP 变量的解释力;
- (2) SOPs 之间的交互作用分析 (SOPs vs SOPs)
 - (a) 不同 SOPs 组合的解释力;
 - (b) 不同变量类别的 SOPs 组合的解释力;
- (3) SOPs 与 Vars 的交互分析 (SOPs vs Vars)
 - (a) 不同 SOPs 与 Vars 组合的解释力;
 - (b) 不同类别的 SOPs 与不同类别的 Vars 组合的解释力;
- (4) 总体变量解释力分析 (Overall SOPs vs Overall Vars)
 - (a) 所有 SOPs 变量的解释力;
 - (b) 所有 Vars 变量的解释力;
 - (c) 所有 SOPs 与 Vars 组合的解释力;
- (5) 验证与比较
 - (a) Vars 的解释力与 SOPs 的解释力的比较;
 - (b) Vars 交互的解释力与 SOPs 交互的解释力的比较;

(c) SOPs, Vars, 与 SOPs + Vars 组合的解释力比较;
更详细的验证和比较结果将在结果分析中展示。

5.3 结果分析

5.3.1 空间异常模式 (SOPs) 的空间分布

首先应用 SDO 模型计算了小麦产量的空间异常模式 (SOPs), 并将其可视化展示。SOPs 的计算是通过在每一个格网点上逐窗口移动计算其周围不同缓冲区半径内的异常信息来实现的。有多少个格网就需要计算多少次, 因此格网数越多, 计算的时间成本越高。

图 5-2 展示了不同缓冲区半径下的降水, 温度, 和 EVI 的 SOPs 分布情况。为了在更小尺度上识别异常信息, 同时防止空间关联性较小的异常信息对结果产生影响, 将缓冲区设置为从 0.1° 到 1° , 缓冲间隔为 0.1° , 一共 10 个缓冲区, 每个缓冲区识别出一定数量的异常信息, 根据这些异常信息的强度及其积极作用和消极作用计算出了每个解释变量对应的 20 个 SOPs (含正负), 小麦产量的解释变量选择了 9 个, 那么就总共得到了 180 组扩展的解释变量, 即 SOPs, 这些变量相对于原始的 9 个变量包含了更丰富的空间信息。

从图 5-2 可以看到, SOPs 随着缓冲区的变大, 其空间模式的强度也在增大, 这是因为更大的缓冲区能够捕捉到更广泛的空间异常信息, 尤其是对于小麦产量这种受多种环境因素影响的变量, 其空间分布模式更加复杂。其中在新南威尔士州 (NSW), 维多利亚州 (VIC), 西澳大利亚州 (WA) 以及昆士兰州 (QLD) 等地的部分区域的 SOPs 分布较为明显, 显示出这些地区的小麦产量受环境因素影响较大。

5.3.2 单个变量及 SOP 的解释力

图 5-3 展示了不同变量组合与类别下对小麦产量空间分布的解释力差异。图中 (a) 和 (b) 分别展示了单个原始变量 (Var) 和单个空间模式变量 (SOP) 的解释力。可以看到, 在单一变量的情况下, 两种类型的变量的解释力差异较小, 其主要影响变量也各不相同。

原始变量对小麦产量解释力较强的包括降水 (TP)、温度 (AT) 和增强植被指数 (EVI), 分别为 0.311, 0.284 和 0.273。而 SOPs 的解释力则更为均衡, 但普遍低于单个原始变量, 其解释力最高的三个变量为 ETa (0.128), PTO (0.121) 和 SLT (0.118)。此结果仅为单一变量的解释力分析, 对空间异质性的量化还将从多个方面进行进一步比较验证。

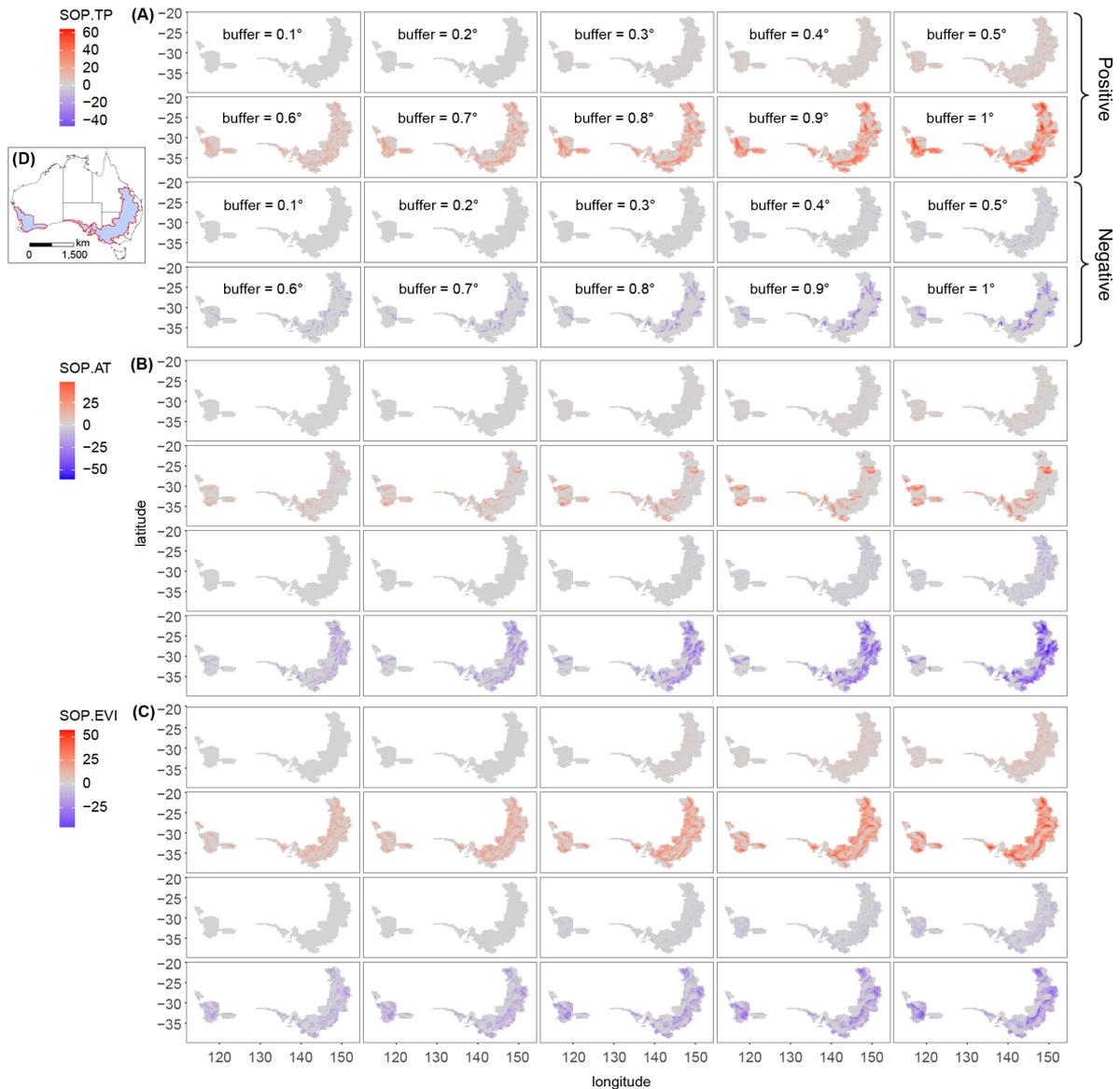


图 5-2 TP、AT 和 EVI 的第二维度空间模式变量 (SOPs) 在不同缓冲区尺度下的空间分布图。图中缓冲区从 0.1° 到 1° ，间隔为 0.1° ，共生成 10 个缓冲区，并分别识别正向和负向 SOPs。

Fig. 5-2 Spatial distributions of second-dimension spatial pattern variables (SOPs) for TP, AT, and EVI across multiple buffer scales. The buffer sizes range from 0.1° to 1° at 0.1° intervals, with both positive and negative SOPs identified within each buffer.

5.3.3 SOPs 间的相互作用机制

图 5-4 展示了 SOPs 之间交互作用的解释力分析，即仅使用不同 SOPs 的组合来计算其空间解释力。图中 (a) 为不同的单个 SOPs 组合下其交互作用的解释力，(b) 为根据解释变量的类别进行分类后的 SOP 分类两两组合后的交互解释力。其中 C1 表示气候变量分类，C2 表示环境变量分类，C3 表示土壤属性变量分类。

表 5-1 列出了小麦产量空间异质性分析中前 10 个 SOPs 相互作用的解释力值。可以很明显地看到，SOPs 之间的交互作用解释力普遍较单一的 SOP 的解释力有着明显

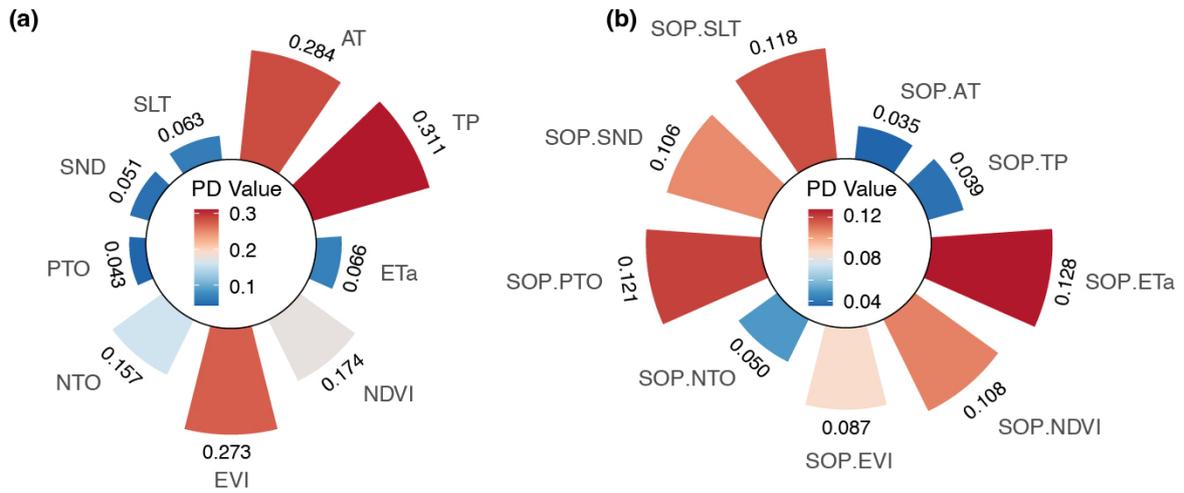


图 5-3 原始变量与空间模式变量解释力对比图。图中：(a) 显示单个原始变量 (Var) 的解释力；(b) 显示单个空间模式变量 (SOP) 的解释力。

Fig. 5-3 Comparison of explanatory power for original variables and spatial pattern variables. Sub-figures: (a) explanatory power of individual original variables (Vars); (b) explanatory power of individual spatial pattern variables (SOPs).

的增加。图 5-4(a) 和表 5-1 中显示，ETa 与 PTO 的组合解释力最高，达到 0.31，其次是 NDVI 与 PTO 的组合 (0.27)，这些两两组合均表现出较强的空间解释力，这些结果表明，SOPs 之间的交互作用能够显著提高对小麦产量空间分布的解释力。

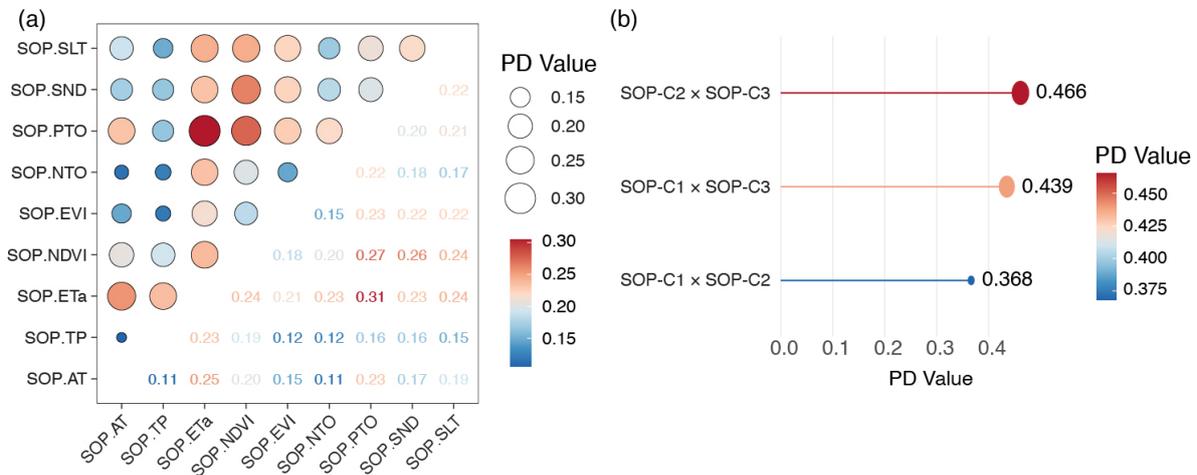


图 5-4 不同 SOPs 组合及分类内两两组合的交互解释力分析图。其中：(a) 为不同 SOPs 组合的交互解释力；(b) 为三类变量中 SOPs 的两两组合解释力。

Fig. 5-4 Interaction explanatory power analysis of SOPs. Subplots include: (a) interaction explanatory power between different SOP combinations; (b) interaction explanatory power of pairwise SOP combinations within three variable categories.

而在不同类别的 SOPs 组合中，环境变量 (C2) 与土壤属性变量 (C3) 的组合解释力较高，达到了 0.466，而气候变量 (C1) 与环境变量 (C2) 的组合解释力最低，为 0.368，但均比单一的原始变量和 SOP 变量都要高。这表明环境和土壤因素对小麦产量

表 5-1 小麦产量空间异质性分析中前 10 个 SOPs 相互作用的解释力值

Tab. 5-1 Power of determinant (PD) values for the top 10 SOPs interactions in the spatial heterogeneity analysis of wheat production

| SOPs | SOPs | <i>qv</i> | <i>Sig.</i> |
|------|------|-----------|------------------------|
| ETa | PTO | 0.305 | 5.07×10^{-10} |
| NDVI | PTO | 0.275 | 5.13×10^{-10} |
| NDVI | SND | 0.263 | 5.85×10^{-10} |
| AT | NDVI | 0.255 | 5.63×10^{-10} |
| NDVI | SLT | 0.243 | 6.74×10^{-10} |
| ETa | SLT | 0.242 | 6.78×10^{-10} |
| ETa | NDVI | 0.237 | 5.65×10^{-10} |
| TP | NDVI | 0.235 | 4.71×10^{-10} |
| ETa | NTO | 0.233 | 7.21×10^{-10} |
| ETa | SND | 0.232 | 9.71×10^{-10} |

的空间分布有更显著的影响。

5.3.4 SOPs 与原始变量的交互作用

上述仅使用了 SOPs 进行交互作用分析，虽然 SOPs 包含了比原始变量更多的空间异常信息。但在实际应用中，SOPs 与原始变量 (Vars) 的组合往往能提供更全面的空间解释力。在本节进一步分析了 SOPs 与原始变量的交互作用。

图 5-5 (a) 展示了单一原始变量与其对应的单一 SOPs 组合后的交互作用解释力。从图中可以明显看出，在单一变量的情况下，加入 SOP 对解释力的变化并不明显，最终的解释力水平仍旧取决于原有样本点位置出的解释变量。表 5-2 列出了小麦产量空间异质性分析中前 30 个 SOPs 与 Vars 相互作用的解释力值。可以看到，SOPs 与原始变量的组合能够显著提高对小麦产量空间分布的解释力。其中，TP 与 SND 的组合解释力最高，达到了 0.317，其次是 TP 与 SLT 的组合 (0.316)，这些结果表明，SOPs 与原始变量的交互作用能够显著提高对小麦产量空间分布的解释力。

图 5-5 (b) 展示了不同类别变量与其对应的 SOPs 组合后的解释力，可以明显看到，气候变量 (C1) 与 SOP 后的土壤属性变量 (SOP-C3) 的组合解释力较高，达到了 0.62，而 SOP 后的气候变量 (SOP-C1) 与环境变量 (C2) 的组合解释力最低，仅为 0.36。这表明气候和土壤因素能够更好的解释小麦产量的空间分布。

图 5-6 (a) 展示了单一原始变量 (Vars) 及单一原始变量与其对应的单一 SOPs 组合 (Vars + SOPs) 后的解释力条形图。而图 5-6 (b) 表示的是这两种组合的解释力差异，可以看到明显看到，在单一变量的情况下，引入 SOPs 并没有对空间解释力有很大的提升，仅 PTO 提升了 0.015，可忽略不计。

图 5-6 (c) 和 (d) 展示了不同类别变量与其对应的 SOPs 组合后的解释力差异。可以

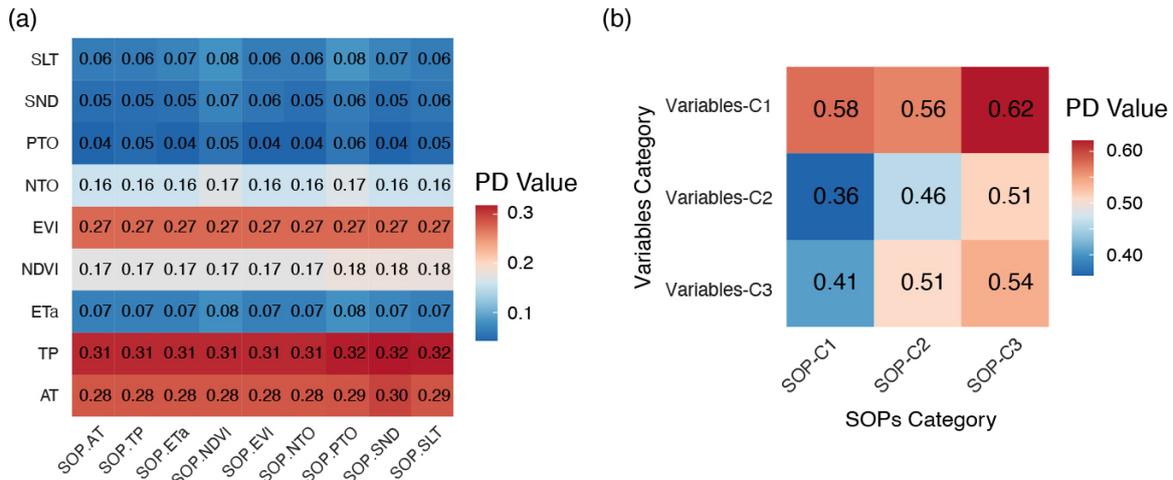


图 5-5 SOP 与原始变量 (Vars) 组合及分类组合的交互解释力分析图。其中：(a) 为 SOP 与 Vars 组合后的交互解释力；(b) 为不同类别下 SOP 与 Vars 组合的交互解释力。

Fig. 5-5 Interaction explanatory power analysis of SOP-Var combinations. Subplots include: (a) interaction explanatory power of SOP-Var combinations; (b) interaction explanatory power of SOP-Var combinations grouped by variable categories.

表 5-2 小麦产量空间异质性分析中前 30 个 SOPs 与 Vars 相互作用的解释力值

Tab. 5-2 Power of Determinant (PD) values for the top 30 SOPs-Vars interactions in the spatial heterogeneity analysis of wheat production

| Vars | SOPs | <i>qv</i> | <i>Sig.</i> | Vars | SOPs | <i>qv</i> | <i>Sig.</i> |
|------|------|-----------|------------------------|------|------|-----------|------------------------|
| TP | SND | 0.317 | 3.37×10^{-10} | TP | SLT | 0.316 | 6.23×10^{-10} |
| TP | PTO | 0.316 | 3.82×10^{-10} | TP | AT | 0.311 | 7.86×10^{-10} |
| TP | TP | 0.311 | 7.86×10^{-10} | TP | ETa | 0.311 | 7.86×10^{-10} |
| TP | NDVI | 0.311 | 7.86×10^{-10} | TP | EVI | 0.311 | 7.86×10^{-10} |
| TP | NTO | 0.311 | 7.86×10^{-10} | AT | SND | 0.296 | 4.49×10^{-10} |
| AT | SLT | 0.286 | 7.75×10^{-10} | AT | PTO | 0.285 | 5.22×10^{-10} |
| AT | NTO | 0.284 | 2.45×10^{-10} | AT | EVI | 0.284 | 2.45×10^{-10} |
| AT | NDVI | 0.284 | 2.45×10^{-10} | AT | ETa | 0.284 | 2.45×10^{-10} |
| AT | TP | 0.284 | 2.45×10^{-10} | AT | AT | 0.284 | 2.45×10^{-10} |
| EVI | AT | 0.273 | 7.28×10^{-10} | EVI | TP | 0.273 | 7.28×10^{-10} |
| EVI | ETa | 0.273 | 7.28×10^{-10} | EVI | NDVI | 0.273 | 7.28×10^{-10} |
| EVI | EVI | 0.273 | 7.28×10^{-10} | EVI | NTO | 0.273 | 7.28×10^{-10} |
| EVI | PTO | 0.273 | 7.28×10^{-10} | EVI | SND | 0.273 | 7.28×10^{-10} |
| EVI | SLT | 0.273 | 7.28×10^{-10} | NDVI | SND | 0.182 | 1.91×10^{-10} |
| NDVI | PTO | 0.182 | 2.63×10^{-10} | NDVI | SLT | 0.180 | 3.89×10^{-10} |

看到，每一种类别的变量在加入 SOP 后的变量之后，其解释力都有所提升，尤其是气候变量 (C1) 的提升幅度更大，达到了 0.156。这也间接可以说明，模型引入的变量数量越多，空间解释力越高，但并不是所有的变量都能对空间分布有显著的影响，尤其是当变量之间存在较强的相关性时，可能会导致解释力的冗余。

5.3.5 模型总体验证

在模型总体验证这一节，通过原始变量，SOPs 变量以及原始变量与 SOPs 变量的组合来比较其对小麦产量空间分布的整体解释力。

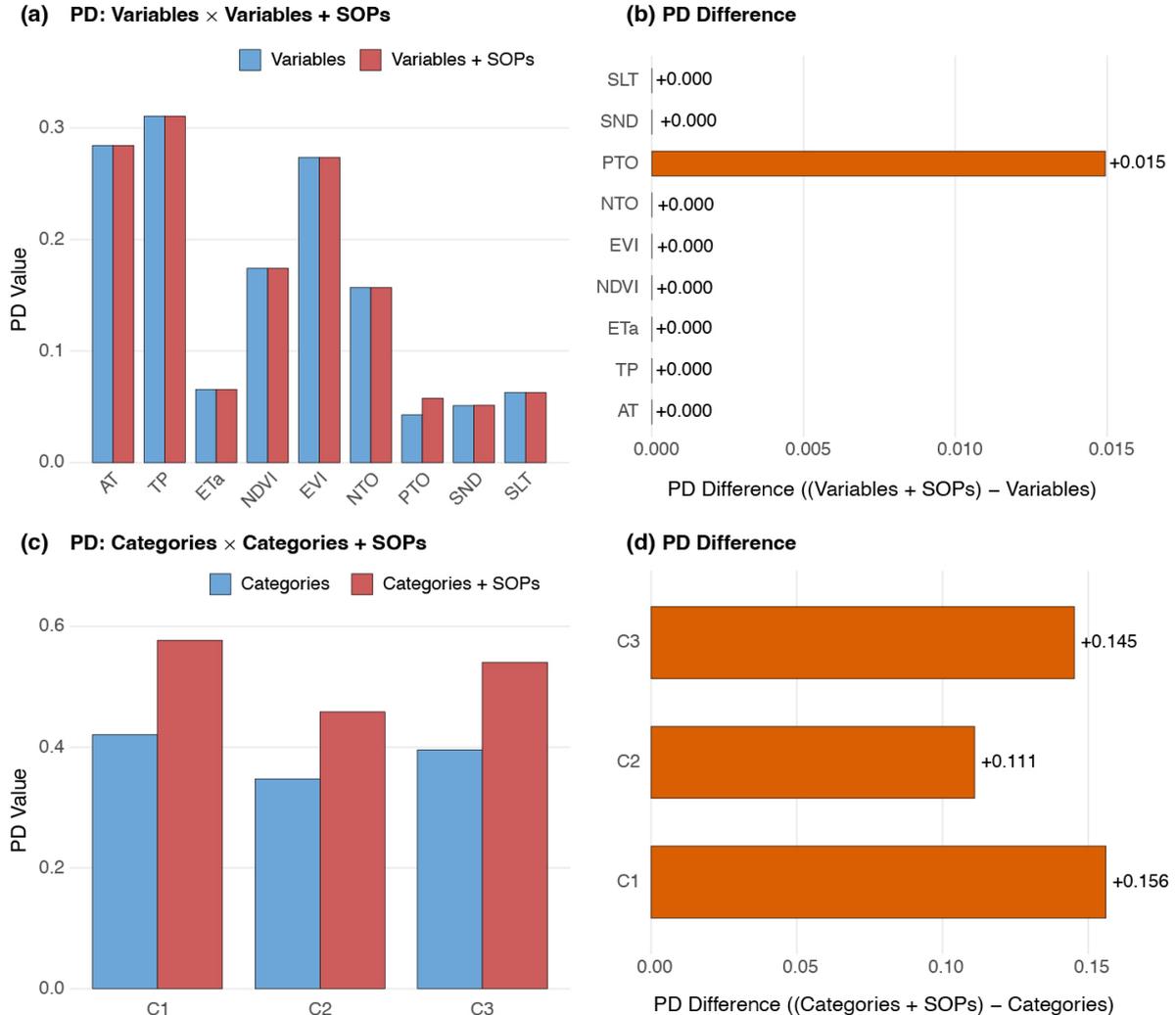


图 5-6 原始变量与其 SOP 组合及不同类别变量组合后的解释力差异图。图中：(a-b) 比较单个原始变量与其对应 SOP 组合后的解释力；(c-d) 展示不同类别变量与其组合 SOP 后的解释力差异。

Fig. 5-6 Comparison of explanatory power differences when combining Vars with SOPs and across variable categories. Subfigures: (a-b) comparison of explanatory power between single Vars and their corresponding SOP-enhanced combinations; (c-d) comparison of explanatory power between variable categories and their respective SOP combinations.

如图 5-7 的条形图所示，A1 表示所有 SOPs 变量的 PD 值，A2 表示 SOPs 与原始变量 (Vars) 组合后的 PD 值，A3 表示所有原始变量的 PD 值，其对应的解释力数值分别为 0.503, 0.688 和 0.635。显然，加入 SOPs 变量后，整体的解释力有了显著的提升，从原始变量的 0.635 提高到了 0.688，约提升了 8.3%，这表明 SOPs 变量能够更好地捕

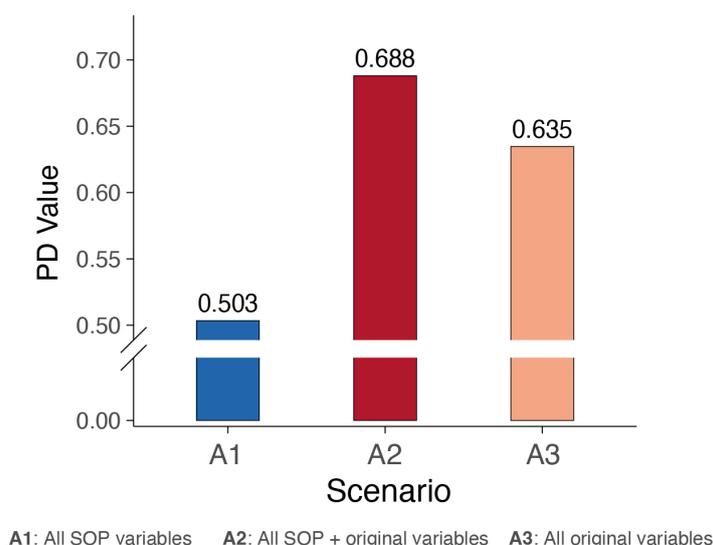


图 5-7 不同变量组合下对小麦产量空间异质性的整体解释力 (PD) 比较图。A1 表示所有 SOPs 变量的 PD 值, A2 表示 SOPs 与原始变量 (Vars) 组合后的 PD 值, A3 表示所有原始变量的 PD 值。

Fig. 5-7 Comparison of overall Power of Determinant (PD) values across different variable combinations for spatial heterogeneity in wheat production. A1 represents the PD values of all SOP variables, A2 represents the PD values of combined SOP and original variables (Vars), and A3 represents the PD values of all original variables.

捉和量化小麦产量的空间分布模式。

但是所有 SOPs 变量整体的解释力则相对较低, 这是因为 SOPs 变量虽然包含了更多的空间异常信息, 但其本身的解释力并不如原始变量强, 且 SOPs 变量是基于原始变量生成的, 没有原始变量的空间信息加成, 其异质性解释力度则较弱。这也说明了在实际应用中, 单一的 SOPs 变量可能无法完全替代原始变量, 而是需要与原始变量结合使用, 以达到更好的空间解释效果。

5.4 扩展分析-量化和解释澳大利亚大麦产量空间分布模式

本小节是继 SOH 模型在澳大利亚小麦中的应用之后, 对大麦产量空间分布模式的进一步分析。通过对大麦产量的空间分布进行 SOH 模型分析, 旨在揭示其空间异质性及影响因素。在大麦产量的研究中, 选择了更多的数据量 (12 种, 4 大类), 以及更小的空间范围, 以期获得更精细的验证与分析。

5.4.1 数据收集与变量选择

大麦是澳大利亚种植与生产的仅次于小麦的主要粮食作物。图 5-8 是澳大利亚大麦的主要产出分布区域, 该数据是从澳大利亚统计局获取的 2023-24 年最新大麦产量数据, 以统计区域等级 2 (Statistical Area Level 2, SA2) 为空间单元, SA2 是比 LGA 范围更小的统计区域等级, 具有更小的尺度和更高的精度。虽然从图中可以很明显看

到西澳地区是大麦的主要产出区，但西澳的空间单元尺度相对较大，数量更少，其数据量和精度不足以支撑验证 SOH 模型。

因此，本节选择东澳大利亚地区的大麦产区作为主要的研究对象，不含西澳大利亚州以及塔斯马尼亚州。东澳大利亚地区的大麦产区主要集中在新南威尔士州 (NSW)、维多利亚州 (VIC)、昆士兰州 (QLD) 和南澳大利亚州 (SA)，这些地区的空间单元更为集中，其数量较多，且数据质量较高，适合进行 SOH 模型的验证与分析。大麦产量的空间单元是包含 187 个 SA2 的统计区域。在每个空间单元上计算大麦产量的平均值，作为目标对象数据。

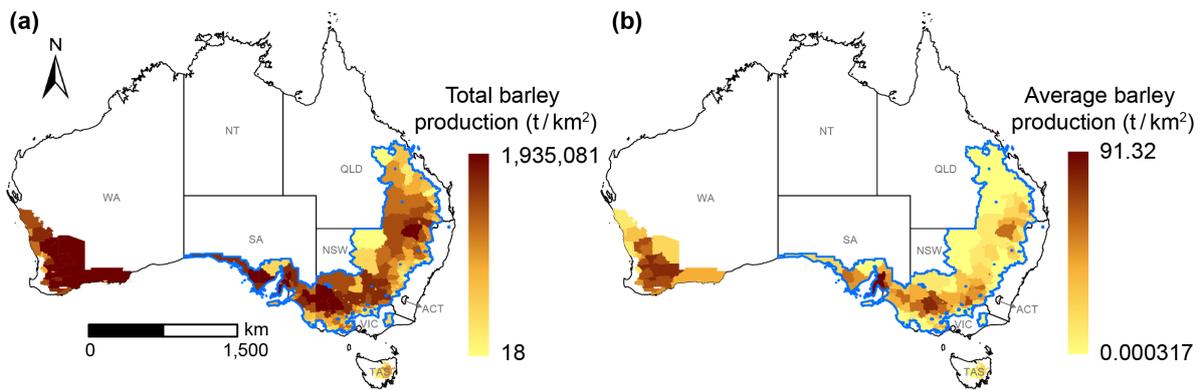


图 5-8 2023-24 年大麦总产量 (a) 与平均产量 (b) 的空间分布图，图中蓝色范围为本研究的研究区域。

Fig. 5-8 Spatial distribution of total (a) and average (b) barley production in 2023-24. The blue shaded area indicates the study region.

影响大麦产量的空间解释变量根据文献和经验，选择了气候、地形、土壤属性、植被等多方面因素。主要以下 12 个变量作为研究对象 (表 5-3)，具体数据来源和预处理方法同小麦产量分析。

这些解释变量同样经过重采样、裁剪、极端值处理和标准化等预处理，以确保数据的质量和一致性。将这些变量的待分析格网数据的空间分辨率统一为 1000 米，以便于后续的空间模式变量 (SOPs) 计算。图 5-9 展示了这些解释变量的空间分布情况，其中 (a) - (d) 为气候变量，包括气温 (AT)、降水量 (TP)、饱和水汽压差 (VPD) 和蒸散量 (ETa); (e) 为地形变量，海拔 (ELE); (f) - (j) 为土壤属性变量，包括土壤有机碳 (SOC)、总氮 (NTO)、总磷 (PTO)、 $\text{pH}(\text{CaCl}_2)$ (pHc) 和有效持水量 (AWC); (k) - (l) 为植被变量，包括 NDVI 和 NPP。

5.4.2 大麦产量的 SOPs 的空间分布

在本节同样应用 SDO 模型逐网格窗口计算了大麦产量的空间异常模式 (SOPs)，并将其可视化展示。图 5-10 是对大麦产量的空间差异性影响最高的三个解释变量 (TP,

表 5-3 潜在影响大麦产量空间分异的解释变量概述

Tab. 5-3 A summary of explanatory variables that potentially affect spatial disparities of barley production.

| 类别 Category | 变量 Variable | 代码 Code | 数据产品 Product | 分辨率 Resolution |
|----------------------------------|--|---|------------------------------|-------------------|
| 气候 Climate | 气温 Air temperature | AT | ERA5_Land | 0.1° |
| | 总降水量 Total precipitation | TP | ERA5_Land | 0.1° |
| | 饱和水汽压差 Vapor pressure deficit | VPD | ERA5_Land | 0.1° |
| | 蒸散量 Evapotranspiration | ETa | TERN/AET/CMRSET Landsat V2.2 | 30 m |
| 地形 Topography | 海拔 Elevation | ELE | DEM-S (Geoscience Australia) | 30 m |
| 土壤属性 Soil Properties | 土壤有机碳 Soil organic carbon | SOC | CSIRO/SLGA | 92.77 m |
| | 总氮 Total nitrogen | NTO | CSIRO/SLGA | 92.77 m |
| | 总磷 Total phosphorus | PTO | CSIRO/SLGA | 92.77 m |
| | pH (CaCl ₂) pH (CaCl ₂) | pHc | CSIRO/SLGA | 92.77 m |
| | 有效持水量 Available water capacity | AWC | CSIRO/SLGA | 92.77 m |
| | 植被 Vegetation | 归一化植被指数 Normalized difference vegetation index | NDVI | MOD13A2 V6.1 |
| 净初级生产力 Net primary production | | NPP | MOD17A3HGF V6.1 | 500 m |

AT, EVI) 所对应的不同缓冲区下的正负第二维度空间模式变量 (SOPs)。

SOPs 的计算方法与小麦产量的 SOPs 计算方法相同, 使用了相同的缓冲区设置, 从 0.1° 到 1°, 缓冲间隔为 0.1°, 一共 10 个缓冲区。每个缓冲区内识别出一定数量的异常信息, 根据这些异常信息的强度及其积极作用和消极作用计算出了每个解释变量对应的 20 个 SOPs (SOPs⁺ 和 SOPs⁻)。SOPs 依旧遵循缓冲区范围越大, SOPs⁺ 和 SOPs⁻ 的数量和强度越高的规律。

以 AT 为例, 可以看到, 随着缓冲区的增大, AT 的 SOPs⁺ 和 SOPs⁻ 的空间分布模式也在不断变化。较小的缓冲区 (如 0.1°) 下, SOPs⁺ 和 SOPs⁻ 的空间分布较为分散, 而随着缓冲区的增大, 其空间分布模式逐渐集中, 显示出更明显的空间异常信息。同时也可以看出识别出来的 AT 异常信息点主要分布在昆士兰州北部和东部, 新南威尔士州东部, 以及维多利亚州南部等区域。这表明 AT 对大麦产量的影响在不同空间尺度上存在差异。

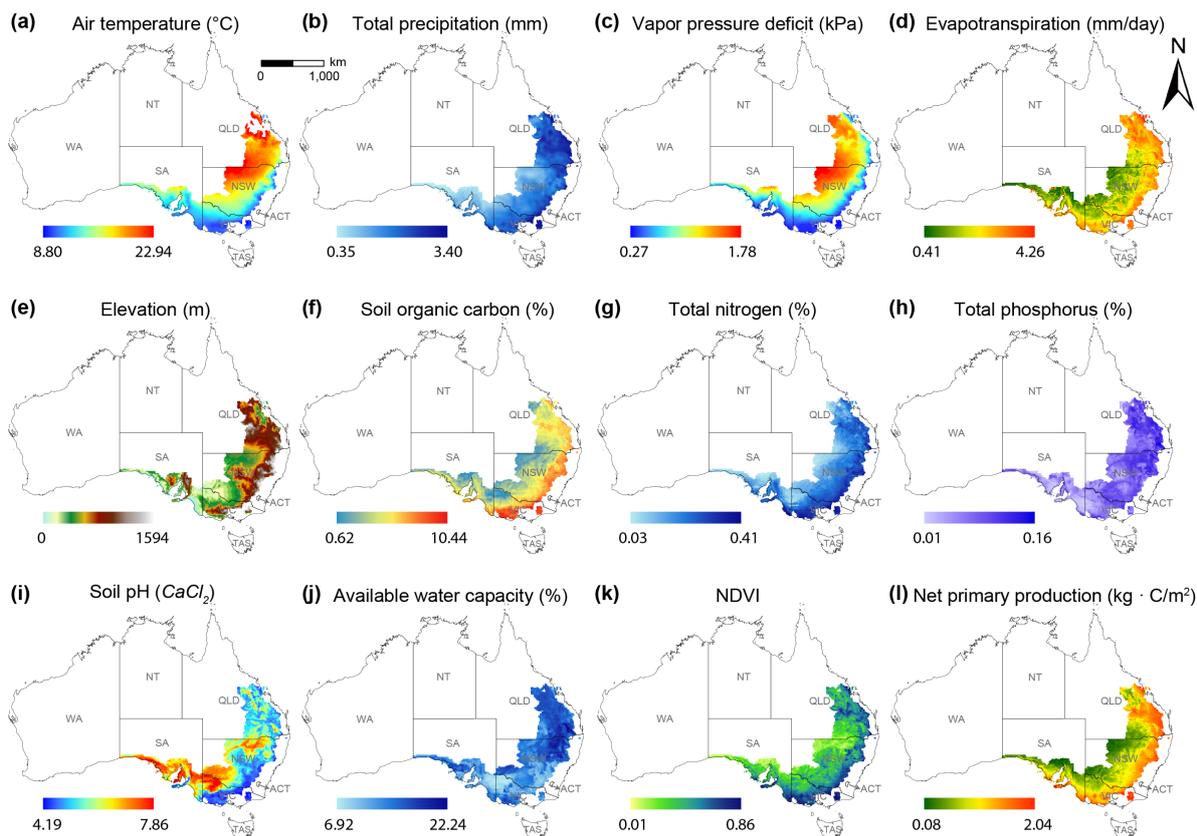


图 5-9 2023-24 年潜在影响澳大利亚大麦产量的解释变量空间分布图。图中包含多个子图，按变量类别分组：(a)–(d) 为气候变量，包括气温 (AT)、降水量 (TP)、饱和水汽压差 (VPD) 和蒸散量 (ETa)；(e) 为地形变量，海拔 (ELE)；(f)–(j) 为土壤属性变量，包括土壤有机碳 (SOC)、总氮 (NTO)、总磷 (PTO)、 $\text{pH}(\text{CaCl}_2)$ (pHc) 和有效持水量 (AWC)；(k)–(l) 为植被变量，包括 NDVI 和 NPP。

Fig. 5-9 Spatial distribution of potential explanatory variables influencing barley production in Australia in 2023-24. The figure includes multiple subplots grouped by category: (a)–(d) are climatic variables, including air temperature (AT), total precipitation (TP), vapor pressure deficit (VPD), and evapotranspiration (ETa); (e) is a topographic variable—elevation (ELE); (f)–(j) are soil property variables, including soil organic carbon (SOC), total nitrogen (NTO), total phosphorus (PTO), $\text{pH}(\text{CaCl}_2)$ (pHc), and available water capacity (AWC); (k)–(l) are vegetation variables, including NDVI and net primary production (NPP).

5.4.3 单个 SOP 与 Var 的解释力分析

SOH 模型在大麦产量案例中的验证方法同样与小麦案例相同。首先分析单个 SOPs 变量和单个原始变量 (Vars) 的解释力，然后分析 SOPs 之间的交互作用，以及 SOPs 与 Vars 不同组合的交互作用，最后比较 SOPs 与原始变量的组合对大麦产量空间分布的整理解释力。

图 5-11 (a)(b) 分别表示了单个原始变量和单个 SOPs 变量的空间解释力差异。可以看到，单个 SOPs 变量的解释力普遍低于单个原始变量，这是因为 SOPs 包含了空间异常信息，但是其原本样本点的空间信息则更依赖原始变量。大麦的验证结果与小麦的验证结果类似。

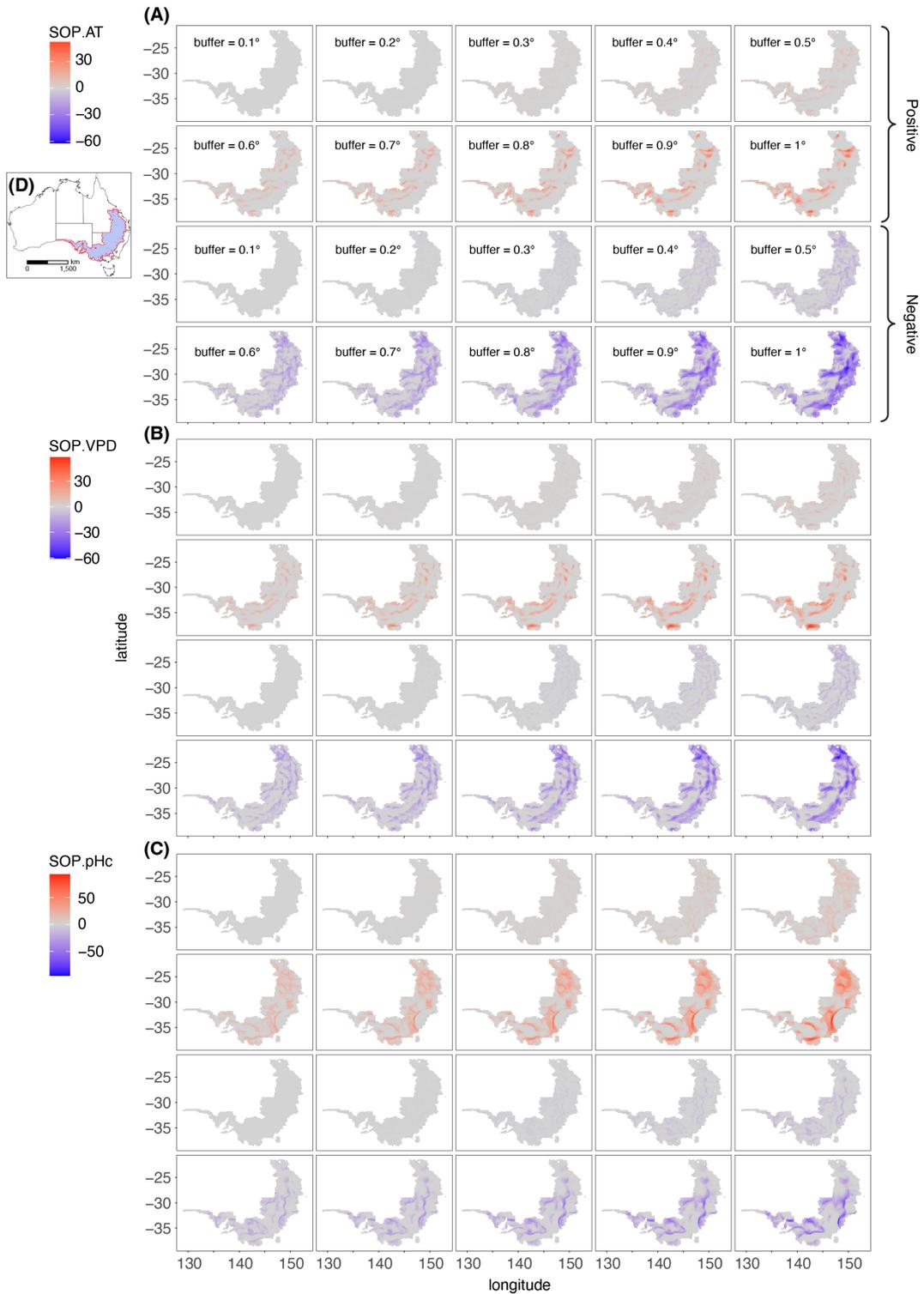


图 5-10 大麦产量空间异质性中解释力 (PD) 最高的三个解释变量的第二维离群模式 (SOPs) 的空间分布。

Fig. 5-10 Spatial distribution of second-dimension outlier patterns (SOPs) for the three explanatory variables with highest determinant power (PD) in barley production spatial heterogeneity.

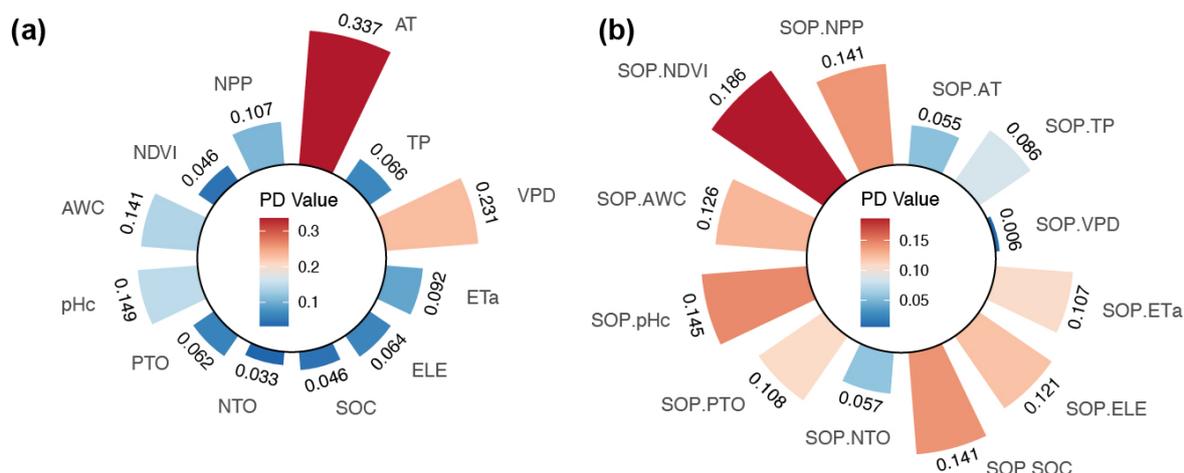


图 5-11 原始变量与空间模式变量解释力对比图 (大麦)。图中：(a) 显示单个原始变量 (Var) 的解释力；(b) 显示单个空间模式变量 (SOP) 的解释力。

Fig. 5-11 Comparison of explanatory power for individual original variables (Vars) and spatial pattern variables (SOPs) in barley production. Subplots: (a) explanatory power of individual original variables (Vars); (b) explanatory power of individual spatial pattern variables (SOPs).

5.4.4 SOPs 间的交互作用分析

针对 SOPs 之间的交互作用分析，类似的比较了不同 SOPs 的组合。单个 SOP 之间的两两组合的解释力分析结果如图 5-12 (a) 所示。很明显，SOPs 之间的交互作用能够显著提高对大麦产量空间分布的解释力。结合表 5-4 中的前 10 个 SOPs 相互作用的解释力值，可以看到，NPP 与 AWC 的组合解释力最高，达到了 0.416，其次是 AT 与 NDVI 的组合 (0.365) 和 pHc 与 AWC 的组合 (0.361)，这些结果表明，单个 SOPs 两两之间的交互作用对空间模式解释力远大于单个原始变量和单个 SOP 变量。

同理，根据解释变量所对应的大类，比较了不同类别的 SOPs 组合的解释力，如图 5-12 (c) 所示。可以看到，气候变量 (C1) 与地形变量 (C2) 的组合的解释力最低，仅为 0.408，而土壤属性变量 (C3) 与植被变量 (C4) 的组合解释力最高，为 0.553。结合小麦产量的验证结果，可以认为土壤属性变量对大麦和小麦产量的空间分布有更显著的影响。这可能由于澳大利亚广袤的土地资源和多样的土壤类型，使得在澳洲这篇大陆上土壤属性对作物生长和产量的影响更大，但不可否认其他要素的关键作用。

5.4.5 SOPs 与 Vars 的交互作用分析

原始变量 (Vars) 是包含样本点位置的空间信息最多的变量，因此将其与 SOPs 结合，计算其交互作用下的解释力尤为必要。

图 5-13 (a) 的热力图表示的是单个原始变量与其对应的单个 SOPs 组合后的交互作用解释力。从图中可以明显看出，在单一变量的情况下，加入 SOP 对解释力的变化并不明显，最终的解释力水平仍旧取决于原有样本点位置出的解释变量。进一步说明了

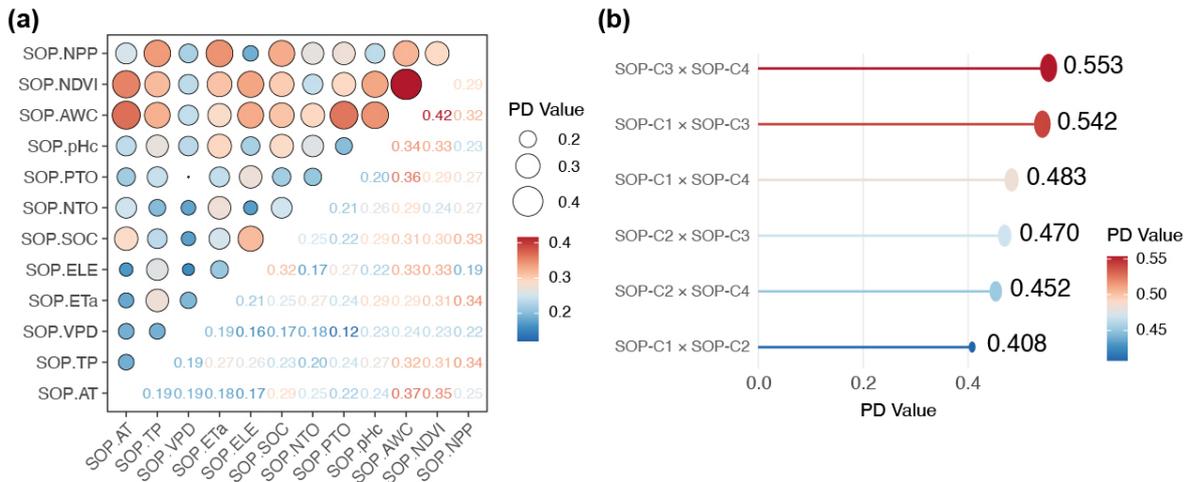


图 5-12 SOPs 组合及分类内两两组合的交互解释力分析图。图中：(a) 为不同 SOPs 组合的交互解释力；(c) 为三类变量中 SOPs 的两两组合解释力。

Fig. 5-12 Interaction explanatory power analysis of SOPs combinations. Subplots include: (a) interaction explanatory power between different SOP combinations; (c) interaction explanatory power of pairwise SOP combinations within four variable categories.

表 5-4 大麦产量空间异质性分析中前 10 个 SOPs 相互作用的解释力值

Tab. 5-4 Power of determinant (PD) values for the top 10 SOPs interactions in the spatial heterogeneity analysis of barley production

| SOPs | SOPs | <i>qv</i> | <i>Sig.</i> |
|------|------|-----------|------------------------|
| NPP | NDVI | 0.416 | 9.46×10^{-10} |
| AT | AWC | 0.365 | 8.67×10^{-10} |
| pHc | AWC | 0.361 | 9.48×10^{-10} |
| AT | NDVI | 0.353 | 6.58×10^{-10} |
| ETa | NPP | 0.343 | 7.53×10^{-10} |
| PTO | AWC | 0.343 | 6.31×10^{-10} |
| TP | NPP | 0.338 | 7.16×10^{-10} |
| SOC | NDVI | 0.331 | 8.89×10^{-10} |
| PTO | NDVI | 0.331 | 9.55×10^{-10} |
| SOC | AWC | 0.327 | 8.45×10^{-10} |

原始变量的重要性，SOPs 变量仅能起到辅助作用。

解释力排前 30 的单一原始变量与 SOP 的两两组合如表 5-5 所列，显然原始的气温 (AT) 对大麦产量的空间异质性影响更大，其所对应的组合都表现出很高的解释力，均大于 0.3。

图 5-13 (b) 是不同类别变量与其对应的 SOPs 组合后的解释力，可以明显看到，原始气候变量 (C1) 所控制的四个组合均表现出极其高的解释力，其与四个 SOP 类别组合所对应的解释力分别为：Vars-C1 + SOP-C1:0.70，Vars-C1 + SOP-C2:0.65，Vars-C1 + SOP-C3:0.73，Vars-C1 + SOP-C4:0.67，远超其他类别的组合。这表明在这个模型验证

表 5-5 大麦产量空间异质性分析中前 30 个 SOPs 与 Vars 相互作用的解释力值

Tab. 5-5 Power of Determinant (PD) values for the top 30 SOPs–Vars interactions in the spatial heterogeneity analysis of barley production

| Vars | SOPs | qv | $Sig.$ | Vars | SOPs | qv | $Sig.$ |
|------|------|-------|------------------------|------|------|-------|------------------------|
| AT | TP | 0.345 | 5.01×10^{-10} | VPD | NDVI | 0.238 | 6.71×10^{-10} |
| AT | NPP | 0.345 | 6.15×10^{-10} | VPD | NTO | 0.236 | 1.05×10^{-10} |
| AT | NDVI | 0.345 | 2.04×10^{-10} | VPD | VPD | 0.231 | 2.41×10^{-10} |
| AT | AT | 0.344 | 4.03×10^{-10} | VPD | ETa | 0.231 | 2.41×10^{-10} |
| AT | ETa | 0.342 | 4.32×10^{-10} | VPD | ELE | 0.231 | 2.41×10^{-10} |
| AT | AWC | 0.342 | 9.02×10^{-10} | VPD | SOC | 0.231 | 2.41×10^{-10} |
| AT | VPD | 0.337 | 1.76×10^{-10} | VPD | PTO | 0.231 | 2.41×10^{-10} |
| AT | ELE | 0.337 | 1.76×10^{-10} | VPD | pHc | 0.231 | 2.41×10^{-10} |
| AT | SOC | 0.337 | 1.76×10^{-10} | VPD | AWC | 0.231 | 2.41×10^{-10} |
| AT | NTO | 0.337 | 1.76×10^{-10} | pHc | TP | 0.161 | 5.88×10^{-10} |
| AT | PTO | 0.337 | 1.76×10^{-10} | pHc | ETa | 0.157 | 4.64×10^{-10} |
| AT | pHc | 0.337 | 1.76×10^{-10} | pHc | NDVI | 0.156 | 4.12×10^{-10} |
| VPD | TP | 0.243 | 3.35×10^{-10} | pHc | AWC | 0.155 | 2.11×10^{-10} |
| VPD | AT | 0.240 | 3.66×10^{-10} | pHc | AT | 0.155 | 3.55×10^{-10} |
| VPD | NPP | 0.239 | 2.26×10^{-10} | pHc | PTO | 0.155 | 9.88×10^{-10} |

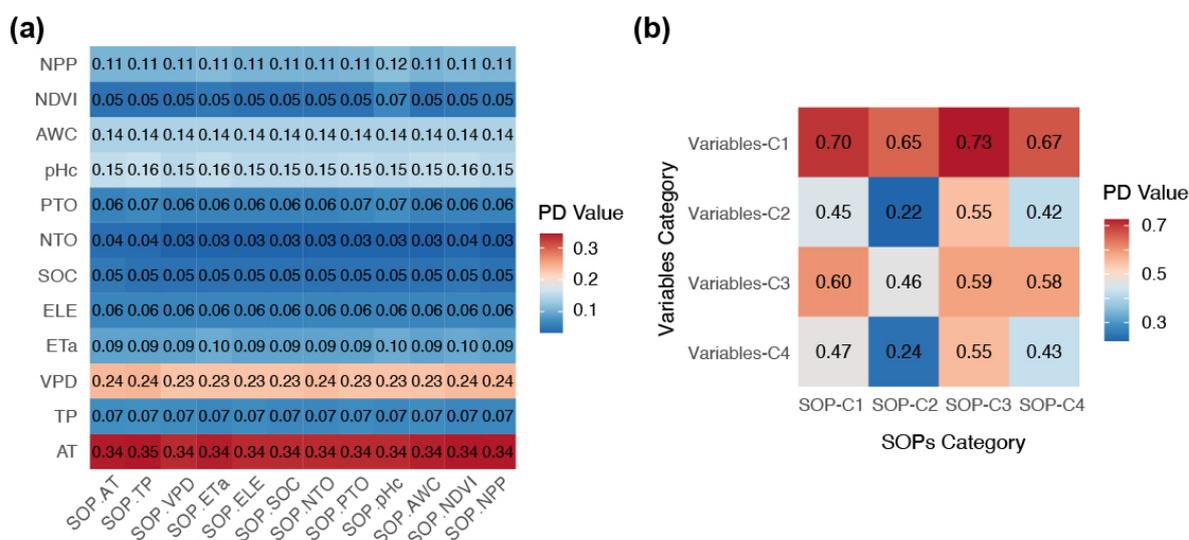


图 5-13 SOP 与原始变量 (Vars) 组合及分类组合的交互解释力分析图。图中: (b) 为 SOP 与 Vars 组合后的交互解释力; (d) 为不同类别下 SOP 与 Vars 组合的交互解释力。

Fig. 5-13 Interaction explanatory power analysis of SOP–Var combinations. Subplots include: (b) interaction explanatory power of SOP–Var combinations; (d) interaction explanatory power of SOP–Var combinations grouped by variable categories.

过程中, 气候因素对大麦产量的空间分布有更显著的影响。

图 5-14 (a) 展示了单一原始变量 (Vars) 与单一原始变量和其对应的单一 SOPs 组合 (Vars + SOPs) 后的解释力条形图。而图 5-14 (b) 表示的是这两种组合的解释力差异,

可以看到明显看到，在单一变量的情况下，引入 SOPs 对空间解释力有微弱的提升，但提升的变量数量较多。这可能归因于本节在大麦案例中使用了更多的变量和更小的空间范围。

在不同类别的组合比较中，从图 5-14 (c) 和 (d) 可以看到，每一种类别的变量在加入 SOP 后的变量之后，其解释力都有明显的提升，尤其是植被变量 (C4) 的提升幅度更大，达到了 0.319。相较于单个变量的组合，分类变量的组合有着更优的表现。

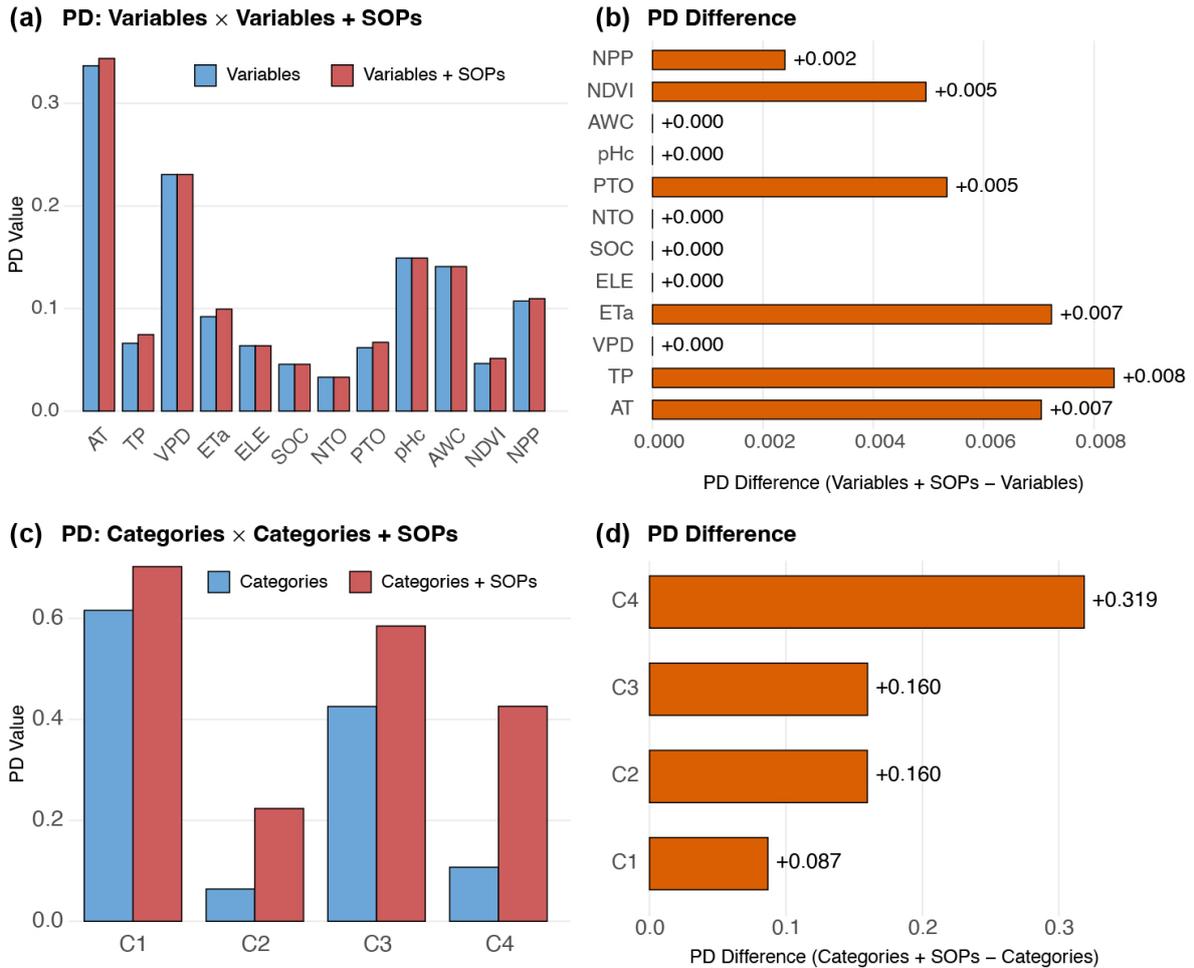


图 5-14 大麦产量空间分布中变量组合与类别下解释力差异图。图中：(a-b) 比较单个原始变量与其对应 SOP 组合后的解释力；(c-d) 展示不同类别变量与其组合 SOP 后的解释力差异。

Fig. 5-14 Comparison of explanatory power differences across variable combinations and categories for barley production spatial distribution. Subplots: (a-b) comparison of explanatory power between individual Vars and their corresponding SOP-enhanced combinations; (c-d) comparison of explanatory power between variable categories and their respective SOP combinations.

5.4.6 SOH 模型整体的验证与比较

同样的，全部 SOPs 变量，全部原始变量组合 SOPs 变量，以及全部原始变量的整体解释力比较如图 5-15 所示。当所有的原始变量与 SOPs 组合后，其所包含的空间信息最多，尤其是对空间异质量化最关键也是最容易忽略的异常信息进行详细的刻画后的变量组合，其空间解释力达到了 0.732，比全部原始变量的解释力的 0.69 提升了 6.1%，虽然提升的幅度不高，但也足以表达出样本点周围的空间异常对于目标变量的空间模式与空间异质性的量化和解释的重要性。

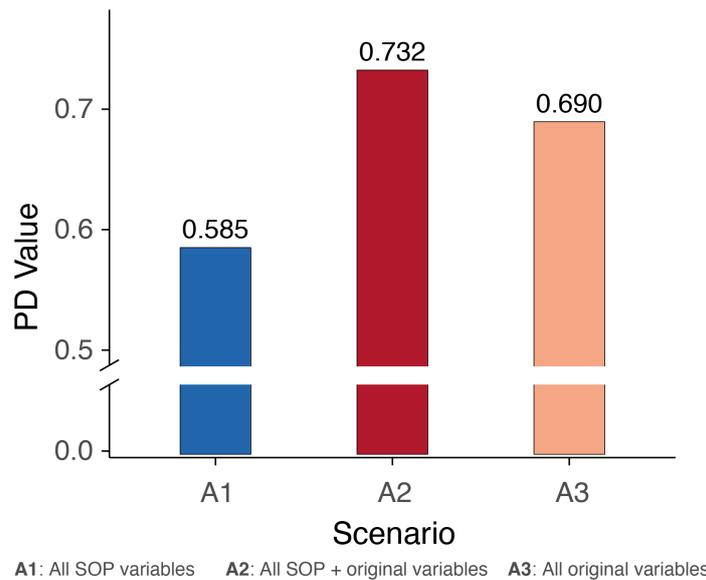


图 5-15 不同变量组合下对大麦产量空间异质性的整体解释力 (PD) 比较图。A1 表示所有 SOPs 变量的 PD 值，A2 表示 SOPs 与原始变量 (Vars) 组合后的 PD 值，A3 表示所有原始变量的 PD 值。

Fig. 5-15 Comparison of overall Power of Determinant (PD) values across different variable combinations for spatial heterogeneity in barley production. A1 represents the PD values of all SOP variables, A2 represents the PD values of combined SOP and original variables (Vars), and A3 represents the PD values of all original variables.

5.4.7 大麦案例的结果分析

通过不同变量组合的空间解释力分析，我们可以得出以下结论：

- (1) 第二维度空间异常对空间异质性的量化和解释有着重要的辅助和促进作用。
- (2) SOPs 变量与原始变量的组合能够显著提高对大麦产量空间分布的解释力。
- (3) 越多的解释变量引入，SOH 模型的空间解释力越高。
- (4) 气候因素和土壤属性对大麦产量的空间分布有更显著的影响，尤其是气温、降水量和有效持水量。

5.5 讨论

本研究基于第二维度异常 (SDO) 模型, 提出了第二维度异常驱动的异质性量化模型 (SOH), 并分别以澳大利亚小麦和大麦产量为案例, 系统评估了原始变量 (Vars)、空间模式变量 (SOPs) 及其组合对作物空间分布异质性的解释力。通过多尺度缓冲区下的正负离群信息提取, 本研究首次将空间局部极值与结构突变信息引入地理探测器框架, 实现了空间模式变量对目标变量空间分布的精准量化和解读。以下针对主要发现及其意义展开讨论。

首先, SOH 模型显著提升了空间分布解释力。以小麦案例为例, 所有原始变量的 PD 值为 0.635, 加入 SOPs 后提升至 0.688, 增幅约 8.3%; 大麦案例中, 从 0.690 提升至 0.732, 增幅约 6.1%。这一结果表明, SOPs 所承载的缓冲区内离群信息有效补充了原始变量的空间信息, 提高了对空间异质性的捕捉能力, 验证了异常模式在空间异质性分析中的价值。

其次, SOPs 之间以及 SOPs 与 Vars 之间的交互作用揭示了要素协同机制。两两 SOPs 组合中, 小麦中 ETa-PTO 组合 PD 最高 (0.305), 大麦中 NPP-NDVI 组合 PD 最高 (0.416), 说明不同类别的异常模式具有协同增效作用; 而 SOPs 与 Vars 组合中, 小麦以 TP-SND (0.317) 和 TP-SLT (0.316) 最为显著, 大麦以 AT-TP (0.345) 和 AT-NDVI (0.345) 效果最佳, 反映出气候变量与土壤或植被异常信息的互补性。可见, 多尺度离群信息与原始样本点位置共同作用, 能够更全面地揭示驱动空间分布异质性的关键因素。

再次, 不同类别变量在 SOH 模型中的表现存在差异。小麦案例中, 环境与土壤属性组合 (C2-C3) 的交互 PD 达到 0.466, 气候与环境组合 (C1-C2) 最低为 0.368; 大麦案例中, 土壤属性与植被组合 (C3-C4) PD 最高达 0.553, 而气候与地形组合 (C1-C2) 最低为 0.408。这表明, 在澳大利亚大陆尺度下, 土壤与植被因素对作物产量空间分布的影响更为显著, 而气候与环境因素的协同作用相对较弱, 提示了不同因子类别在区域管理和模型构建中的优先级。

另一方面, 案例对比揭示了作物类型的响应差异。尽管小麦与大麦均受气候、土壤、植被等多因素影响, 但大麦对植被生产力 (NPP) 和土壤水分保持 (AWC) 的异常模式更为敏感, 小麦则更强调蒸散量 (ETa) 与总磷 (PTO) 等因子的空间异常。这种差异可能源于两种作物的生长周期、水分需求及农艺管理方式不同, 为农业决策提供了针对性指导。

本研究在理论层面丰富了空间异质性分析方法。传统地理探测器侧重原始变量分类与方差解释, 缺乏对空间局部极值及结构突变的量化; SOH 模型通过提取多尺度离群模式, 构建空间模式变量, 拓展了空间解释变量的内涵, 为复杂地理过程提供了新

的建模思路。此外，将决策树与地理探测器结合，实现了变量选择与解释力评估的一体化，为后续模型优化提供了可复制的框架。

在实践层面，SOH 模型具有重要应用价值。对于农业管理机构，可基于 SOPs 识别高风险或潜在高产区域，指导精准灌溉、施肥和种植结构调整；对于政策制定者，可量化不同环境与土壤政策的区域效应，为优化农业补贴与资源配置提供数据支撑；对于地理信息系统开发者，可将 SOPs 框架集成到空间决策支持系统，实现实时监测与预警。

然而，本研究亦存在若干局限。第一，缓冲区半径与间隔设置需平衡空间分辨率与计算成本，本研究以 0.1° – 1° 为例，未来可结合自适应算法优化；第二，SOPs 计算基于静态年际数据，未考虑时间序列动态变化，难以捕捉季节性或极端事件的瞬时效应；第三，数据质量与尺度匹配影响结果可靠性，不同数据源精度差异需在跨区域应用时加以校准；第四，当前模型侧重单因子与双因子交互，尚未深入挖掘多因子高阶交互机制。

针对上述不足，后续研究可从以下方面展开：一是将 SOH 模型应用于其他作物（如大豆、水稻）或生态系统（如森林、草地），检验模型的普适性与可扩展性；二是引入时空序列数据，构建时变 SOPs，揭示不同生育期与极端气候下的异常模式；三是融合遥感高光谱与无人机航测数据，提升空间分辨率与变量多样性；四是结合深度学习与图模型，探索高阶空间依赖与因果路径，为复杂地理系统建模提供新方法；五是开展敏感性分析与不确定性评估，明确模型参数对结果的影响范围，增强决策可靠性。

综上所述，SOH 模型通过量化多尺度空间离群信息，实现了对地理现象空间异质性的深入解读，为空间分析理论与实践提供了有益补充。未来在数据、算法与应用场景层面的进一步完善，将推动 SOH 模型在更加广泛的地理、环境与农业领域中发挥更大作用。

5.6 小结

本章基于第二维度异常的概念，结合决策树和地理探测器方法，提出了第二维度异常驱动的异质性量化模型（SOH），并将其应用于小麦和大麦产量的空间分布模式分析。通过对小麦和大麦产量的空间异质性进行了多种不同形式的验证和分析，本章证明了 SOPs 模型在空间数据分析中的有效性和实用性，其与原始解释变量的结合提高了空间异质性的量化能力。

第六章 结论与展望

6.1 主要结论

本论文以澳大利亚粮食主产区为研究对象，系统开展了三阶段研究：首先，构建融合空间自相关—空间分层异质性—决策树的地理空间机器学习框架，对小麦产量的空间差异及主导因子进行定量识别；其次，引入“第二维度异常（SDO）”理念，将传统模型最易忽略的局地空间异常信息显式融入特征体系，显著提升空间预测精度；最后，基于 SDO 概念，结合决策树与地理探测器提出了“第二维度异常驱动的异质量化模型（SOH）”，并分别在小麦与大麦两条作物链上验证其解释力和实用性。三部分研究层层递进，共同构建了“空间模式识别—预测精度提升—空间异质量化”的完整技术路线，为精准农业、区域种植管理与可持续粮食安全提供了方法论与实证支撑。本研究的主要结论如下：

（1）识别小麦产量的空间差异与决定因素

通过构建融合空间自相关、空间分层异质性与决策树的地理空间机器学习模型，揭示澳大利亚小麦产量的空间差异及其主导因素。结果表明，小麦产量在空间上呈现显著的分异格局，主要集中在新南威尔士州（NSW）南部、南澳大利亚州（SA）南部，和西澳大利亚州（WA）西部与南部。通过地理最优区域分层异质性模型（GOZH），将澳大利亚的小麦划分为四个地理最优区域，最大化区域间的差异并最小化区域内的产量差异。同时发现 ASP, EVI, NTO 等影响因子对小麦产量的影响最大，为区域化种植管理与可持续农业发展提供量化依据。

（2）第二维度异常（SDO）提升空间预测精度

第二维度异常（SDO）方法通过引入样本点邻域的空间异常结构信息，构建跨越属性维度与空间结构维度的“第二维度”变量体系，并将这些异常信息编码为空间模式变量（SOPs）与原始环境因子联合建模。在澳大利亚小麦产量预测中，SDO-SVM 模型相较非空间 SVM 模型的判定系数 R^2 提高 33.9%，均方根误差（RMSE）与平均绝对误差（MAE）分别降低 20.9% 和 19.6%，显著提升预测精度并减少误差。同时，敏感性分析显示，在缓冲间隔、最大缓冲阈值及异常值判据等参数变化下，交叉验证 RMSE（CV-RMSE）的相对波动始终控制在 $\pm 5\%$ 范围内，表明模型具有良好的鲁棒性与可推广性。与传统非空间预测方案相比，SDO 能在保留原始极端值的前提下有效利用其空间上下文信息，展现出更高的可靠性与实用价值。

（3）提出空间异质量化模型 SOH 并将其应用于小麦和大麦案例研究中

本研究基于第二维度异常思想，集成决策树与地理探测器构建了空间异质量化

模型 (SOH)，通过在原始环境因子 (Vars) 基础上引入第二维度异常空间模式变量 (Spatial Outlier Patterns, SOPs)，显著提升了作物产量空间分布的解释力与异质性刻画能力。以澳大利亚小麦为例，Vars+SOPs 组合的整体解释力 (PD) 达到 0.688，较仅使用原始变量提升 8.3%，并揭示了局部异常格局与环境因子耦合的多尺度驱动机制。进一步在 2023-24 年澳大利亚大麦数据上开展跨作物扩展与多形式模型验证，在更精细的空间尺度与更丰富的解释变量集下，SOH 模型整体 PD 达到 0.732，相比原始变量基准模型提高 6.1%。结果表明，SOH 框架具有较强的稳健性、迁移性与泛化潜力，可为多粮种精细化管理、区域尺度农业监测与可持续决策提供方法支撑。

6.2 创新点

本研究将地理空间分析方法和农作物产量进行结合，通过识别影响澳大利亚小麦产量的决定因素以及确定其地理最优产区，分析小麦产量的空间差异性，基于此，提出了第二维度异常空间预测方法和第二维度异常值驱动的异质性量化模型，为空间预测和异质性量化解释提供了新的思路和方法，也为农作物的管理和决策提供了科学依据。主要创新点包括：

(1) 融合式地理空间机器学习框架

首次将空间自相关分析、空间分层异质性和决策树算法有机整合，形成自动化“最优区划 + 决定因素识别”流程，显著提升了对作物产量空间差异的解析深度与分区决策的可操作性。

(2) “第二维度异常”(SDO) 概念与特征生成机制

创新性地提出区别于传统邻域统计的第二维度异常方法，利用样本点周边的空间异常结构信息，构建全新的变量体系，并与原有样本点数据进行结合，从而提高空间预测的精度。经过不同角度的验证，以及与传统的非空间预测方法进行比对，证明了该方法在空间预测中的有效性和可靠性。

(3) SDO 驱动的异质性量化模型 (SOH)

创新性地将决策树离散化与地理探测器解释力评估嵌入 SDO 体系，构建 SOH 模型，实现了单因子、交互因子及跨类别因子的异质性量化；并在小麦与大麦双作物上验证了模型的广泛适用性与迁移能力。

6.3 研究不足与展望

本研究仍然存在很多不足之处，希望在未来的研究工作中可以加以改进。

(1) 由于澳大利亚统计局仅在地方政府区 (Local Government Area, LGA) 尺度发布与全套环境协变量配对的小麦产量数据，且数据完全统一，本文聚焦于 2015-16 与

2020–21 两个农业普查年份。虽然单季数据一般足以进行地理探测器等模型的分析，我们使用的两年数据对比分析为 GOZH 框架的稳健性提供了额外验证。然而，LGA 尺度仍相对粗糙，未来应将小麦种植区下采样至 1 km 甚至 100 m 栅格，以更精准地捕获区内变异。当更长且一致的时间序列数据可用时，可在时空框架内应用 GOZH，分析年际动态并验证本研究识别的空间决定因子的持久性。近期发展出的 Robust GOZH (Luo et al. 2025) 在区域划分过程中引入 Q 函数以优化复杂度参数 (Complexity Parameter, CP)，为 CP 的数据驱动选择提供了途径并增强了分区稳定性。未来研究将评估该方法，以进一步提升区域划分结果的性能与可解释性。

(2) 本研究是将每个 LGA 的平均小麦产量与气候、土壤和环境变量建立关系，而这些解释变量是基于整个行政边界区域平均得到的。这种空间尺度的不匹配可能引发“生态谬误”与“可变区域单元问题” (Modifiable Areal Unit Problem, MAUP)。为尽量减少这类偏差，研究在变量聚合之前，首先将每个环境变量图层限制在 LGA 范围内已知的小麦种植带区域，以确保响应变量与解释变量在空间上具有可比性。尽管如此，部分聚合误差仍不可避免。因此，本研究结果更应被视为对 SDO 模型有效性与潜力的数值性验证，而非可直接应用于实际决策的产品。未来，若能获取田块尺度或高分辨率的格网产量数据，模型可直接迁移至更精细的分辨率，并结合高分辨率分析、多尺度敏感性测试、播种面积加权与空间显式建模等方法，进一步缓解生态谬误与 MAUP 所带来的影响。

(3) 本研究构建的模型框架虽然在解释小麦及大麦的产量空间差异性方面表现出良好性能，但由于大麦与小麦在生理生态特性及栽培管理上存在较高相似性，当前研究成果主要适用于旱作农业体系中的禾谷类作物。模型对于其他种类作物 (如水稻、玉米、豆类等) 在不同农业系统 (如灌溉农业、热带种植园) 中的解释力与迁移适用性尚未经过充分验证，这限制了模型在更广泛作物类型和农业区域中的普适性应用。在后续研究中，我们将重点评估该模型框架在不同生态区、不同作物类型中的适应性与稳健性，通过引入了更多样的作物产量数据及生长环境参数，进一步明确其适用边界与优势应用场景，从而增强该地理空间建模方法在支持全球多种粮食作物可持续生产与决策中的实际价值。

(4) 在未来的工作中，我们可以进一步挖掘第二维度空间特征与异常类型 (如第二维度异质性或复杂度)，以完善现有的空间统计与机器学习模型。SDO 方法具有跨学科推广潜力，可为农业、气象、生态、城市研究及绿地健康等领域提供更准确、可靠的空间预测支持。

参考文献

- 郝展宏, 叶松林, 蔡东玉, 宁鹏, 张丽娟, 米国华. 2025. 华北地区冬小麦—夏玉米绿色吨半粮田生产技术. *现代农业科技*, (12): 202~205.
- 刘文毅, 梁楠鑫, 杜虹, 刘中杰, 辛梓豪, 喻姝研, 张娅. 2025. 基于地理加权回归模型的豫西熊耳山地区金矿地球化学异常提取研究. *黄金*, 46(06): 62~68.
- 刘相荣. 2025. 浅析小麦-玉米轮作高产栽培技术. *河北农业*, (06): 64~65.
- 时运佳, 袁宏伟, 朱梅, 曹秀清, 刘硕硕. N.d. 淮北平原旱涝急转条件下冬小麦减产规律研究. *节水灌溉*, 1~12.
- 童葶. 2008. 澳大利亚粮食市场开放与中国粮食安全. [硕士学位论文]. 南京农业大学.
- 王劲峰, 徐成东. 2017. 地理探测器: 原理与展望. *地理学报*, 72(1), 116: 116~134. <https://www.geog.com.cn/CN/10.11821/dlxb201701010>.
- 韦加固, 杨佳庆, 冉莉萍, 董颀闻, 吴雨阳, 熊飞, 余徐润. 2025. 小麦穗发育机制及影响因素的研究进展. *麦类作物学报*, 45(08): 1121~1135.
- 魏煜轩, 吴一平, 刘利花. 2025. 传统农区新型城镇化水平的区域差异及空间收敛性. *河南农业大学学报*, 1~20.
- 吴久江. 2024. 华北平原冬小麦净初级生产力对干旱的响应机制及未来气候变化应对策略研究. [博士学位论文]. 西北农林科技大学.
- 杨福. 2025. 现代农业技术在小麦生产中的应用与前景展望. *种子科技*, 43(10): 222~224.
- 张绍华, 凌彬, 祁攀, 王双雄, 孔令达. 2024. 澳大利亚进口小麦质量分析与应用. *粮食加工*, 49(06): 9~12.
- 张幸福, 姜元俊, 阿比尔的. N.d. 考虑空间异质性的降雨滑坡易发性预测研究. *工程科学与技术*, 1~17.
- 朱行. 2016. 澳大利亚小麦的生产加工与出口. *现代面粉工业*, 30(01): 42~43.
- Abakar K A, Yu C. 2014. Performance of SVM based on PUK kernel in comparison to SVM based on RBF kernel in prediction of yarn tenacity. *Indian Journal of Fibre & Textile Research*, 39(1).
- Ahrends H E, Eugster W, Gaiser T, Rueda-Ayala V, Hüging H, Ewert F, Siebert S. 2018. Genetic yield gains of winter wheat in Germany over more than 100 years (1895–2007) under contrasting fertilizer applications. *Environmental Research Letters*, 13(10): 104003.
- Ajami M, Heidari A, Khormali F, Zeraatpisheh M, Gorji M, Ayoubi S. 2020. Spatial variability of rainfed wheat production under the influence of topography and soil properties in loess-derived soils, northern Iran. *International Journal of Plant Production*, 14: 597–608.
- Ale-Kajbaf F, Khadem-Alali M, Marashi S, Sheidai M, Noormohammadi Z. 2024. Spatial analysis and sequential equation modeling (SEM) of genetic diversity in date palm (*Phoenix dactylifera* L.) *The Nucleus*,

- 1–12.
- Anselin L. 1995. Local indicators of spatial association—LISA. *Geographical analysis*, 27(2): 93–115.
- Anselin L. 2005. Exploring spatial data with GeoDaTM: a workbook. *Center for spatially integrated social science*, 1963: 157.
- Anselin L. 2019. Quantile local spatial autocorrelation. *Letters in Spatial and Resource Sciences*, 12(2): 155–166.
- Araki S, Shimadera H, Yamamoto K, Kondo A. 2017. Effect of spatial outliers on the regression modelling of air pollutant concentrations: A case study in Japan. *Atmospheric Environment*, 153: 83–93.
- Australian Bureau of Agricultural and Resource Economics and Sciences. 2025. Dashboard Visualization. <https://public.tableau.com/app/profile/australian.bureau.of.agricultural.and.resource.economics.and.sci/viz/shared/C9ZTTSS4C>. Accessed: 2024-04-03.
- Australian Bureau of Statistics. 2017. Agricultural Commodities, Australia, 2015-16 financial year. <https://www.abs.gov.au/AUSSTATS/abs%40.nsf/DetailsPage/7121.02015-16?OpenDocument>. Dataset. Accessed: 2024-05-17.
- Australian Bureau of Statistics. 2020. Australian Statistical Geography Standard (ASGS): Volume 3 - Non ABS Structures, June 2020. <https://www.abs.gov.au/AUSSTATS/abs%40.nsf/DetailsPage/1270.0.55.0.03June%202020?OpenDocument>. Dataset. Accessed: 2024-05-17.
- Australian Bureau of Statistics. 2021a. Australian Statistical Geography Standard (ASGS) Edition 3, Digital boundary files. <https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/access-and-downloads/digital-boundary-files>. Dataset. Accessed: 2024-05-17.
- Australian Bureau of Statistics. 2021b. Value of Agricultural Commodities Produced by Local Government Area - 2020-21. https://public.tableau.com/app/profile/australian.bureau.of.agricultural.and.resource.economics.and.sci/viz/AgCensus_LGA_2020-21/StoryLGA. Dataset. Accessed: 2024-05-17.
- Australian Bureau of Statistics. 2022. Agricultural Commodities, Australia, 2020-21 financial year. <https://www.abs.gov.au/statistics/industry/agriculture/agricultural-commodities-australia/2020-21>. Dataset. Accessed: 2024-05-17.
- Australian Export Grains Innovation Centre. 2022. Australian wheat information brochure. <https://www.aegic.org.au/australian-grains/wheat/>. Report. Accessed: 2024-05-17.
- Australian Export Grains Innovation Centre. 2025. Australian wheat information brochure. <https://www.aegic.org.au/grains/wheat/>. Accessed: 2025-04-03.
- Awad M, Khanna R. 2015. Support vector regression. In: *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*. Springer: 67–80.
- Baba A M, Midi H, Abd Rahman N H. 2022. Spatial Outlier Accommodation Using a Spatial Variance Shift Outlier Model. *Mathematics*, 10(17): 3182.
- Bai H, Wang H, Li D, Ge Y. 2023. Information consistency-based measures for spatial stratified hetero-

- geneity. *Annals of the American Association of Geographers*, 113(10): 2512–2524.
- Basso B, Fiorentino C, Cammarano D, Cafiero G, Dardanelli J. 2012. Analysis of rainfall distribution on spatial and temporal patterns of wheat yield in Mediterranean environment. *European Journal of Agronomy*, 41: 52–65.
- Belgiu M, Drăguț L. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114: 24–31.
- Berke O. 2001. Modified median polish kriging and its application to the Wolfcamp–Aquifer data. *Environmetrics: The official journal of the International Environmetrics Society*, 12(8): 731–748.
- Bivand R S, Wong D W. 2018. Comparing implementations of global and local indicators of spatial association. *Test*, 27(3): 716–748.
- Breiman L. 1984. *Classification and Regression Trees* (1st ed.) Routledge: 368.
- Brereton R G, Lloyd G R. 2010. Support vector machines for classification and regression. *Analyst*, 135(2): 230–267.
- Brisson N, Gate P, Gouache D, Charmet G, Oury F.-X, Huard F. 2010. Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *Field crops research*, 119(1): 201–212.
- Brokamp C, Rao M, Ryan P, Jandarov R. 2017. A comparison of resampling and recursive partitioning methods in random forest for estimating the asymptotic variance using the infinitesimal jackknife. *stat*, 6(1): 360–372.
- Bureau of Meteorology. 2022. Annual Climate Statement 2021. Australian Government Bureau of Meteorologyurl+urldate.
- Cai J, Hu W, Yang Y, Yan H, Chen F. 2024. Outlier detection in spatial error models using modified thresholding-based iterative procedure for outlier detection approach. *BMC Medical Research Methodology*, 24(1): 89.
- Cammarano D, Payero J, Basso B, Wilkens P, Grace P. 2012. Agronomic and economic evaluation of irrigation strategies on cotton lint yield in Australia. *Crop and pasture science*, 63(7): 647–655.
- Chen Y, Zhang Z, Tao F, Wang P, Wei X. 2017. Spatio-temporal patterns of winter wheat yield potential and yield gap during the past three decades in North China. *Field Crops Research*, 206: 11–20.
- Cherkassky V, Ma Y. 2004. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks*, 17(1): 113–126.
- Chu L, Huang C, Liu Q, Cai C, Liu G. 2019. Spatial heterogeneity of winter wheat yield and its determinants in the Yellow River Delta, China. *Sustainability*, 12(1): 135.
- Cliff A D, Ord K. 1970. Spatial autocorrelation: a review of existing and new measures with applications. *Economic Geography*, 46(sup1): 269–292.
- Constantinidou K, Hadjinicolaou P, Zittis G, Lelieveld J. 2016. Effects of climate change on the yield of winter wheat in the eastern Mediterranean and Middle East. *Climate Research*, 69(2): 129–141.
- Department of Agriculture, Fisheries and Forestry. 2023. Australian wheat exports. <https://www.agricultur>

- e.gov.au/about/news/australian-wheat-export-snapshot-jan-23. Accessed: 2024-04-03.
- Diacono M, Castrignanò A, Troccoli A, De Benedetto D, Basso B, Rubino P. 2012. Spatial and temporal variability of wheat grain yield and quality in a Mediterranean environment: A multivariate geostatistical approach. *Field Crops Research*, 131: 49–62.
- Didan K. 2021. MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V061. *distributed by NASA EOSDIS Land Processes Distributed Active Archive Center*, <https://doi.org/10.5067/MODIS/MOD13A2.061> [2024-05-17].
- Din S U, Yamamoto K. 2024. Urban Spatial Dynamics and Geo-informatics Prediction of Karachi from 1990–2050 Using Remote Sensing and CA-ANN Simulation. *Earth Systems and Environment*, 1–20.
- Ding X, Liu J, Yang F, Cao J. 2021. Random radial basis function kernel-based support vector machine. *Journal of the Franklin Institute*, 358(18): 10121–10140.
- Ding Y, Zhang M, Qian X, Li C, Chen S, Wang W. 2019. Using the geographical detector technique to explore the impact of socioeconomic factors on PM_{2.5} concentrations in China. *Journal of cleaner production*, 211: 1480–1490.
- Dormann C F, McPherson J M, Araújo M B, Bivand R, Bolliger J, Carl G, Davies R G, Hirzel A, Jetz W, Kissling W D, et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 609–628.
- Drucker H, Burges C J, Kaufman L, Smola A, Vapnik V. 1996. Support vector regression machines. *Advances in neural information processing systems*, 9.
- Duncan E G, O’Sullivan C A, Roper M M, Peoples M B, Treble K, Whisson K. 2017. Crop and microbial responses to the nitrification inhibitor 3, 4-dimethylpyrazole phosphate (DMPP) in Mediterranean wheat-cropping systems. *Soil research*, 55(6): 553–566.
- Evgeniou T, Pontil M. 1999. Support vector machines: Theory and applications. In: *Advanced course on artificial intelligence*. Springer: 249–257.
- Fan D, Yang F, Pan Z, Su X, Pan Y, Han G, Wang J, Wu D, Dong Z. 2018. Development of an improved model to evaluate vulnerability in spring wheat under climate change in Inner Mongolia. *Sustainability*, 10(12): 4581.
- 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, et al. 2022. Increasing dominance of Indian Ocean variability impacts Australian wheat yields. *Nature Food*, 3(10): 862–870.
- Fletcher A L, Chen C, Ota N, Lawes R A, Oliver Y M. 2020. Has historic climate change affected the spatial distribution of water-limited wheat yield across Western Australia? *Climatic Change*, 159(3): 347–364.
- Fletcher J. 2007. What is heterogeneity and is it important? *Bmj*, 334(7584): 94–96.
- Food and Agriculture Organization of the United Nations. 2019. Crop Prospects and Food Situation - Quarterly Global Report no. 3, September 2019. <https://openknowledge.fao.org/server/api/core/bitstreams/5c492292-232a-4d2e-a015-f7f810709451/content>.

- Fotheringham A S, Yang W, Kang W. 2017. Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, 107(6): 1247–1265.
- Fu Y, Huang J, Shen Y, Liu S, Huang Y, Dong J, Han W, Ye T, Zhao W, Yuan W. 2021. A satellite-based method for national winter wheat yield estimating in China. *Remote Sensing*, 13(22): 4680.
- Georganos S, Grippa T, Niang Gadiaga A, Linard C, Lennert M, Vanhuyse S, Mboga N, Wolff E, Kalogirou S. 2021. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, 36(2): 121–136.
- Georganos S, Kalogirou S. 2022. A forest of forests: a spatially weighted and computationally efficient formulation of geographical random forests. *ISPRS International Journal of Geo-Information*, 11(9): 471.
- Geoscience Australia. 2015. Digital Elevation Model (DEM) of Australia derived from LiDAR 5 Metre Grid. https://developers.google.com/earth-engine/datasets/catalog/AU_GA_DEM_1SEC_v10_DEM-S. Dataset. Accessed: 2024-05-17.
- Gomez-Macpherson H, Richards R. 1995. Effect of sowing time on yield and agronomic characteristics of wheat in south-eastern Australia. *Australian Journal of Agricultural Research*, 46(7): 1381–1399.
- Guerschman J P, McVicar T R, Vleeshower J, Van Niel T G, Peña-Arancibia J L, Chen Y. 2022. Estimating actual evapotranspiration at field-to-continent scales by calibrating the CMRSET algorithm with MODIS, VIIRS, Landsat and Sentinel-2 data. *Journal of Hydrology*, 605: 127318.
- Han J, Liu Y, Hou C, Li J, Wang J, Zhang Q, Yang Q, Chen X, Wu J. 2020. A 1Ns disomic addition from *Psathyrostachys Huashanica* Keng confers resistance to powdery mildew in wheat. *Agronomy*, 10(2): 312.
- Han Y, Wang J, Qiao D. 2024. Assessing climate effects on wheat yield heterogeneity in the North China Plain and evolution from 1960 to 2020. *The Crop Journal*, 12(6): 1791–1800.
- Harris P, Fotheringham A, Crespo R, Charlton M. 2010. The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. *Mathematical Geosciences*, 42: 657–680.
- He Y, Fang J, Xu W, Shi P. 2022. Substantial increase of compound droughts and heatwaves in wheat growing seasons worldwide. *International Journal of Climatology*, 42(10): 5038–5054.
- Hearst M A, Dumais S T, Osuna E, Platt J, Scholkopf B. 1998. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4): 18–28.
- Heng L, Asseng S, Mejahed K, Rusan M. 2007. Optimizing wheat productivity in two rain-fed environments of the West Asia–North Africa region using a simulation model. *European Journal of Agronomy*, 26(2): 121–129.
- Henzell T. 2007. Australian agriculture: its history and challenges. CSIRO publishing.
- Hernandez-Ochoa I M, Asseng S, Kassie B T, Xiong W, Robertson R, Pequeno D N L, Sonder K, Reynolds M, Babar M A, Milan A M, et al. 2018. Climate change impact on Mexico wheat production. *Agricultural*

- and *Forest Meteorology*, 263: 373–387.
- Hersbach H, Bell B, Berrisford P, Biavati G, Horányi A, Muñoz Sabater J, Nicolas J, Peubey C, Radu R, Rozum I, et al. 2023. ERA5 monthly averaged data on single levels from 1979 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*,
- Hodson D, White J. 2007. Use of spatial analyses for global characterization of wheat-based production systems. *The Journal of Agricultural Science*, 145(2): 115.
- Hou Q, Hu Z, Yang Y, Fan X, Zhou J, Zhang L, Duan Y. 2023. The Influence Mechanism of Farmer Behavior on the Spatial Pattern Evolution of Agricultural Production in the Yanhe River Basin. *Land*, 12(7): 1281.
- Howden S, Crimp S, Nelson R. 2010. Australian agriculture in a climate of change. In: *Managing climate change: papers from the Greenhouse 2009 conference*: 101–111.
- Hu J, Song Y, Zhang T. 2025. A local indicator of stratified power. *International Journal of Geographical Information Science*, 39(4): 925–943.
- Hu Y, Yao Y, Kou Z. 2020. Exploring on the climate regionalization of Qinling-Daba mountains based on Geodetector-SVM model. *Plos one*, 15(11): e0241047.
- Hughes N, Donoghoe M, Whittle L. 2020. Farm Level Effects of On-Farm Irrigation Infrastructure Programs in the Southern Murray–Darling Basin. *Australian Economic Review*, 53(4): 494–516.
- Hui J, Qin L, et al. 2019. The impacts of climate change on wheat yield in the Huang-Huai-Hai Plain of China using DSSAT-CERES-Wheat model under different climate scenarios. *Journal of Integrative Agriculture*, 18(6): 1379–1391.
- Jalali H, Van Nieuwenhuysse I, Picheny V. 2017. Comparison of kriging-based algorithms for simulation optimization with heterogeneous noise. *European Journal of Operational Research*, 261(1): 279–301.
- Jeong S G, Karthik G M, Kim E S, Zargar A, Ahn S Y, Sagong M J, Kang S H, Cho J.-W, Kim H S. 2022. Architected heterogeneous alloys with selective laser melting. *Scripta Materialia*, 208: 114332.
- Jia P, Liu S, Yang S. 2023. Innovations in public health surveillance for emerging infections. *Annual Review of Public Health*, 44(1): 55–74.
- Jiang Z. 2018. A survey on spatial prediction methods. *IEEE transactions on knowledge and Data Engineering*, 31(9): 1645–1664.
- Jin N, Tao B, Ren W, He L, Zhang D, Wang D, Yu Q. 2022. Assimilating remote sensing data into a crop model improves winter wheat yield estimation based on regional irrigation data. *Agricultural Water Management*, 266: 107583.
- Jiu-jiang W, Nan W, Hong-zheng S, Xiao-yi M. 2022. Spatial–temporal variation of climate and its impact on winter wheat production in Guanzhong Plain, China. *Computers and Electronics in Agriculture*, 195: 106820.
- Kanso F. 2010. Effect of root zone temperature and changes in precipitation on yield components and yield formation in spring wheat (*Triticum aestivum* L. cv. Triso). PhD thesis. University of Hohenheim.
- Kavitha S, Varuna S, Ramya R. 2016. A comparative analysis on linear regression and support vector re-

- gression. In: 2016 online international conference on green engineering and technologies (IC-GET): 1–5.
- Kelejian H H, Prucha I R. 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of econometrics*, 104(2): 219–257.
- Kim H.-S, Chung C.-K, Kim H.-K. 2016. Geo-spatial data integration for subsurface stratification of dam site with outlier analyses. *Environmental Earth Sciences*, 75: 1–10.
- Kim Y, Kim D, Park J, Jun C. 2024. An Effective Algorithm of Outlier Correction in Space–Time Radar Rainfall Data Based on the Iterative Localized Analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–16.
- Kingwell R. 2020. The Changing Trade Landscape in Asian Grain Markets: An Australian Perspective. *Cereal Foods World*, 65(5).
- Kitchen N, Drummond S, Lund E, Sudduth K, Buchleiter G. 2003. Soil electrical conductivity and topography related to yield for three contrasting soil–crop systems. *Agronomy journal*, 95(3): 483–495.
- Kolasa J, Rollo C D. 1991. Introduction: the heterogeneity of heterogeneity: a glossary. In: Ecological heterogeneity. Springer: 1–23.
- Kravchenko A N, Snapp S S, Robertson G P. 2017. Field-scale experiments reveal persistent yield gaps in low-input and organic cropping systems. *Proceedings of the National Academy of Sciences*, 114(5): 926–931.
- Le Gouis J, Oury F.-X, Charmet G. 2020. How changes in climate and agricultural practices influenced wheat production in Western Europe. *Journal of Cereal Science*, 93: 102960.
- Li H, Calder C A, Cressie N. 2007. Beyond Moran’s I: testing for spatial dependence based on the spatial autoregressive model. *Geographical analysis*, 39(4): 357–375.
- Li S, Zhou C, Wang S, Gao S, Liu Z. 2019. Spatial heterogeneity in the determinants of urban form: an analysis of Chinese cities with a GWR approach. *Sustainability*, 11(2): 479.
- Liu J, Zuo B, Vroman P, Rabenasolo B, Zeng X, Bai L. 2010. Visual quality recognition of nonwovens using wavelet texture analysis and robust Bayesian neural network. *Textile Research Journal*, 80(13): 1278–1289.
- Liu X, Chen F, Lu C.-T. 2014. On detecting spatial categorical outliers. *GeoInformatica*, 18: 501–536.
- Liu Y, Wang Y, Zhang J. 2012. New machine learning algorithm: Random forest. In: International conference on information computing and applications: 246–252.
- Liu Z, Lei H, Sheng H, Wang Y. 2023. Analysis of soil organic matter influencing factors in the Huangshui River Basin by using the optimal parameter-based geographical detector model. *Geocarto International*, 38(1): 2246935.
- Lu Q, Chen F, Hancock K. 2009. On path anomaly detection in a large transportation network. *Computers, Environment and Urban Systems*, 33(6): 448–462.
- Lu W, Atkinson D E, Newlands N K. 2017. ENSO climate risk: predicting crop yield variability and coherence using cluster-based PCA. *Modeling Earth Systems and Environment*, 3: 1343–1359.

- Luo H, Luo P, Meng L. 2025. A Robust Geographically Optimal Zones-based heterogeneity model for analyzing the spatial determinants of national traffic accidents. *GIScience & Remote Sensing*, 62(1): 2448283.
- Luo P, Song Y, Huang X, Ma H, Liu J, Yao Y, Meng L. 2022. Identifying determinants of spatio-temporal disparities in soil moisture of the Northern Hemisphere using a geographically optimal zones-based heterogeneity model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 185: 111–128.
- Luo Q, Bryan B, Bellotti W, Williams M. 2005. Spatial analysis of environmental change impacts on wheat production in Mid-Lower North, South Australia. *Climatic Change*, 72(1-2): 213–228.
- Lv X, Zhang H, Zhang Y, Li J. 2023. Spatial Correlation Evolution and Driving Factors of Wheat Production in China Based on Social Network Analysis. *Agronomy*, 13(2): 515.
- Ma Y, Guo G. 2014. Support vector machines applications: vol. 649. Springer.
- Mao Y, Liu Y, Zhuo L, Wang W, Li M, Feng B, Wu P. 2021. Quantitative evaluation of spatial scale effects on regional water footprint in crop production. *Resources, Conservation and Recycling*, 173: 105709.
- Militino A, Palacios M, Ugarte M. 2006. Outliers detection in multivariate spatial linear models. *Journal of statistical planning and inference*, 136(1): 125–146.
- Millar N, Urrea A, Kahmark K, Shcherbak I, Robertson G P, Ortiz-Monasterio I. 2018. Nitrous oxide (N₂O) flux responds exponentially to nitrogen fertilizer in irrigated wheat in the Yaqui Valley, Mexico. *Agriculture, ecosystems & environment*, 261: 125–132.
- Miller M P, Singer M J, Nielsen D R. 1988. Spatial variability of wheat yield and soil properties on complex hills. *Soil Science Society of America Journal*, 52(4): 1133–1141.
- Min M, Zhao W, Hu T, Chen J, Nie X. 2014. Influential factors of spatial distribution of wheat yield in China during 1978–2007: A spatial econometric analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(11): 4453–4460.
- Mishra U, Drewniak B, Jastrow J D, Matamala R M. 2017. Spatial representation of organic carbon and active-layer thickness of high latitude soils in CMIP5 earth system models. *Geoderma*, 300: 55–63.
- Moran P A. 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2): 17–23.
- Myles A J, Feudale R N, Liu Y, Woody N A, Brown S D. 2004. An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 18(6): 275–285.
- Nabiollahi K, Heshmat E, Mosavi A, Kerry R, Zeraatpisheh M, Taghizadeh-Mehrjardi R. 2020. Assessing the influence of soil quality on rainfed wheat yield. *Agriculture*, 10(10): 469.
- Navada A, Ansari A N, Patil S, Sonkamble B A. 2011. Overview of use of decision tree algorithms in machine learning. In: 2011 IEEE control and system graduate research colloquium: 37–42.
- Nirel R, Mugglestone M A, Barnett V. 1998. Outlier-robust spectral estimation for spatial lattice processes. *Communications in Statistics-Theory and Methods*, 27(12): 3095–3111.
- Obanor F, Neate S, Simpfendorfer S, Sabburg R, Wilson P, Chakraborty S. 2013. *Fusarium graminearum* and *Fusarium pseudograminearum* caused the 2010 head blight epidemics in Australia. *Plant Pathology*,

- 62(1): 79–91.
- Olmstead A L, Rhode P W. 2011. Adapting North American wheat production to climatic challenges, 1839–2009. *Proceedings of the National Academy of sciences*, 108(2): 480–485.
- Orton T G, Mallawaarachchi T, Pringle M J, Menzies N W, Dalal R C, Kopittke P M, Searle R, Hochman Z, Dang Y P. 2018. Quantifying the economic impact of soil constraints on Australian agriculture: A case-study of wheat. *Land Degradation & Development*, 29(11): 3866–3875.
- Overmars K d, De Koning G, Veldkamp A. 2003. Spatial autocorrelation in multi-scale land use models. *Ecological modelling*, 164(2-3): 257–270.
- Pang A, Chang M W L, Chen Y. 2022. Evaluation of random forests (RF) for regional and local-scale wheat yield prediction in Southeast Australia. *Sensors*, 22(3): 717.
- Peeters A, Zude M, Käthner J, Ünlü M, Kanber R, Hetzroni A, Gebbers R, Ben-Gal A. 2015. Getis–Ord’s hot-and cold-spot statistics as a basis for multivariate spatial clustering of orchard tree data. *Computers and Electronics in Agriculture*, 111: 140–150.
- Peng W, Kuang T, Tao S. 2019. Quantifying influences of natural factors on vegetation NDVI changes based on geographical detector in Sichuan, western China. *Journal of cleaner production*, 233: 353–367.
- Petitti D B. 2001. Approaches to heterogeneity in meta-analysis. *Statistics in medicine*, 20(23): 3625–3633.
- Porwollik V, Müller C, Elliott J, Chryssanthacopoulos J, Iizumi T, Ray D K, Ruane A C, Arneth A, Balkovič J, Ciaï P, et al. 2017. Spatial and temporal uncertainty of crop yield aggregations. *European Journal of Agronomy*, 88: 10–21.
- Potgieter A B, Hammer G L, Butler D. 2002. Spatial and temporal patterns in Australian wheat yield and their relationship with ENSO. *Australian Journal of Agricultural Research*, 53(1): 77–89.
- Qader S H, Dash J, Atkinson P M. 2018. Forecasting wheat and barley crop production in arid and semi-arid regions using remotely sensed primary productivity and crop phenology: A case study in Iraq. *Science of the total Environment*, 613: 250–262.
- Qi M, Dixit K, Marshall J D, Zhang W, Hankey S. 2022. National land use regression model for NO₂ using street view imagery and satellite observations. *Environmental Science & Technology*, 56(18): 13499–13509.
- Qiao P, Yang S, Lei M, Chen T, Dong N. 2019. Quantitative analysis of the factors influencing spatial distribution of soil heavy metals based on geographical detector. *Science of the Total Environment*, 664: 392–413.
- Ren K, Song Y, Li L, Mancini F, Xiao Z, Zhang X, Qu R, Yu Q. 2025. Identifying climate and environmental determinants of spatial disparities in wheat production using a geospatial machine learning model. *GIScience & Remote Sensing*, 62(1): 2533487.
- Richter G, Beblík A, Schmalstieg K, Richter O. 1998. N-dynamics and nitrate leaching under rotational and continuous set-aside—a case study at the field and catchment scale. *Agriculture, ecosystems & environment*, 68(1-2): 125–138.

- Rigatti S J. 2017. Random forest. *Journal of insurance medicine*, 47(1): 31–39.
- Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas M. 2015. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore geology reviews*, 71: 804–818.
- Rossel R V, Chen C, Grundy M, Searle R, Clifford D, Campbell P. 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. *Soil Research*, 53(8): 845–864.
- Running S, Zhao M. 2021. MODIS/Terra net primary production gap-filled yearly L4 global 500m SIN grid V061. distributed by NASA EOSDIS Land Processes Distributed Active Archive Center., <https://doi.org/10.5067/MODIS/MOD17A3HGF.061> [2024-05-17].
- Salman H A, Kalakech A, Steiti A. 2024. Random forest algorithm overview. *Babylonian Journal of Machine Learning*, 2024: 69–79.
- Sardà-Palomera F, Brotons L, Villero D, Sierdsema H, Newson S E, Jiguet F. 2012. Mapping from heterogeneous biodiversity monitoring data sources. *Biodiversity and Conservation*, 21(11): 2927–2948.
- Sbahi M, Ziboon A.-R, Hassoon K. 2018. Sentinel-2 imageries with meteorological data utilize remote sensing and GIS for wheat crop prove production. In: MATEC Web of Conferences: vol. 162: 03019.
- Schillerberg T A, Tian D, 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.
- Segal M R. 2004. Machine learning benchmarks and random forest regression.
- Sen Z. 2016. Spatial modeling principles in earth sciences: vol. 10. Springer.
- Sharma R C, Morgounov A I, Braun H J, Akin B, Keser M, Bedoshvili D, Bagci A, Martius C, van Ginkel M. 2010. Identifying high yielding stable winter wheat genotypes for irrigated environments in Central and West Asia. *Euphytica*, 171(1): 53–64.
- Shen J, Evans F H. 2021. The potential of Landsat NDVI sequences to explain wheat yield variation in fields in Western Australia. *Remote Sensing*, 13(11): 2202.
- Shen X, Bao W, Qu K. 2021. Subspace-based preprocessing module for fast hyperspectral endmember selection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 3386–3402.
- Shi Y, Gong J, Deng M, Yang X, Xu F. 2018. A graph-based approach for detecting spatial cross-outliers from two types of spatial point events. *Computers, Environment and Urban Systems*, 72: 88–103.
- Shi Y, Lau A K.-H, Ng E, Ho H.-C, Bilal M. 2021. A multiscale land use regression approach for estimating intraurban spatial variability of PM_{2.5} concentration by integrating multisource datasets. *International journal of environmental research and public health*, 19(1): 321.
- Simões M V, Peterson A T. 2018. Utility and limitations of climate-matching approaches in detecting different types of spatial errors in biodiversity data. *Insect Conservation and Diversity*, 11(5): 407–414.
- Song Y.-Y, Lu Y. 2015. Decision tree methods: applications for classification and prediction. *Shanghai*

- archives of psychiatry*,
- Song Y, Linderholm H W, Wang C, Tian J, Huo Z, Gao P, Song Y, Guo A. 2019. The influence of excess precipitation on winter wheat under climate change in China from 1961 to 2017. *Science of the Total Environment*, 690: 189–196.
- Song Y. 2022. The second dimension of spatial association. *International Journal of Applied Earth Observation and Geoinformation*, 111: 102834.
- Song Y. 2023. Geographically optimal similarity. *Mathematical Geosciences*, 55(3): 295–320.
- Song Y, Wang J, Ge Y, Xu C. 2020. An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: Cases with different types of spatial data. *GIScience & Remote Sensing*, 57(5): 593–610.
- Strandberg J, Sjöstedt de Luna S, Mateu J. 2019. Prediction of spatial functional random processes: comparing functional and spatio-temporal kriging approaches. *Stochastic Environmental Research and Risk Assessment*, 33: 1699–1719.
- Sun S, Yang X, Lin X, Sassenrath G F, Li K. 2018. Winter wheat yield gaps and patterns in China. *Agronomy Journal*, 110(1): 319–330.
- Sun X.-L, Wu Y.-J, Zhang C, Wang H.-L. 2019. Performance of median kriging with robust estimators of the variogram in outlier identification and spatial prediction for soil pollution at a field scale. *Science of the Total Environment*, 666: 902–914.
- Suthaharan S. 2016a. Decision tree learning. In: Machine learning models and algorithms for big data classification: thinking with examples for effective learning. Springer: 237–269.
- Suthaharan S. 2016b. Support vector machine. In: Machine learning models and algorithms for big data classification: thinking with examples for effective learning. Springer: 207–235.
- Taheri Shahraiyani H, Sodoudi S. 2016. Statistical modeling approaches for PM10 prediction in urban areas; A review of 21st-century studies. *Atmosphere*, 7(2): 15.
- Tan X, Guo P.-T, Wu W, Li M.-F, Liu H.-B. 2017. Prediction of soil properties by using geographically weighted regression at a regional scale. *Soil Research*, 55(4): 318–331.
- Tang Y, Wang M, Yang Y, Lan Z, Yue Y. 2023. Robust Large-Scale Collaborative Localization Based on Semantic Submaps With Extreme Outliers. *IEEE/ASME Transactions on Mechatronics*,
- Therneau T, Atkinson B. 2023. rpart: Recursive Partitioning and Regression Trees. [url+urldate](#).
- Thompson E S, Saveyn P, Declercq M, Meert J, Guida V, Eads C D, Robles E S, Britton M M. 2018. Characterisation of heterogeneity and spatial autocorrelation in phase separating mixtures using Moran's I. *Journal of colloid and interface science*, 513: 180–187.
- Tian Y, Wu Z, Li M, Wang B, Zhang X. 2022. Forest fire spread monitoring and vegetation dynamics detection based on multi-source remote sensing images. *Remote Sensing*, 14(18): 4431.
- Troy T J, Kipgen C, Pal I. 2015. The impact of climate extremes and irrigation on US crop yields. *Environmental Research Letters*, 10(5): 054013.

- Ulloa-Espindola R, Perez-Albert Y. 2022. Validation of an urban growth prediction model in Quito (Ecuador) built by using weights of evidence and cellular automata. *EURE-REVISTA LATINOAMERICANA DE ESTUDIOS URBANO REGIONALES*, 48(144).
- Vicente-Serrano S M, Domínguez-Castro F, Reig F, Tomas-Burguera M, Peña-Angulo D, Latorre B, Beguería S, Rabanaque I, Noguera I, Lorenzo-Lacruz J, et al. 2023. A global drought monitoring system and dataset based on ERA5 reanalysis: A focus on crop-growing regions. *Geoscience Data Journal*, 10(4): 505–518.
- Wadoux A M.-C, Padarian J, Minasny B. 2019. Multi-source data integration for soil mapping using deep learning. *Soil*, 5(1): 107–119.
- Wagner H H, Chávez-Pesqueira M, Forester B R. 2017. Spatial detection of outlier loci with Moran eigenvector maps. *Molecular ecology resources*, 17(6): 1122–1135.
- Walker G. 1989. Weather sensitivity of western Canada wheat yield, 1900–1988. *Canadian Journal of Soil Science*, 69(4): 857–865.
- Wang B, Chen C, Li Liu D, Asseng S, Yu Q, Yang X. 2015. Effects of climate trends and variability on wheat yield variability in eastern Australia. *Climate Research*, 64(2): 173–186.
- Wang B, Liu D L, O’Leary G J, Asseng S, Macadam I, Lines-Kelly R, Yang X, Clark A, Crean J, Sides T, et al. 2018. Australian wheat production expected to decrease by the late 21st century. *Global change biology*, 24(6): 2403–2415.
- Wang E, Xu J, Jiang Q, Austin J. 2009. Assessing the spatial impact of climate on wheat productivity and the potential value of climate forecasts at a regional level. *Theoretical and applied climatology*, 95: 311–330.
- Wang J F, Ge Y, Li L F, Meng B, Wu J, Bo Y, Du S, Liao Y, Hu M, Xu C. 2014. Spatiotemporal data analysis in geography. *Acta Geographica Sinica*, 69(9): 1326–1345.
- Wang J F, Li X H, Christakos G, Liao Y L, Zhang T, Gu X, Zheng X Y. 2010. Geographical Detectors-Based Health Risk Assessment and its Application in the Neural Tube Defects Study of the Heshun Region, China. *International Journal of Geographical Information Science*, 24(1): 107–127.
- Wang J, Haining R, Zhang T, Xu C, Hu M, Yin Q, Li L, Zhou C, Li G, Chen H. 2024. Statistical modeling of spatially stratified heterogeneous data. *Annals of the American Association of Geographers*, 114(3): 499–519.
- Wang J, Xu C. 2017. Geodetector: Principle and prospective. *Acta Geographica Sinica*, 72(1), 116: 116.
- Wang J, Chen Q, Chen Y. 2004. RBF kernel based support vector machine with universal approximation and its application. In: International symposium on neural networks: 512–517.
- Wang Y, Xu X, Huang L, Yang G, Fan L, Wei P, Chen G. 2019. An improved CASA model for estimating winter wheat yield from remote sensing images. *Remote Sensing*, 11(9): 1088.
- Wassenaar T, Lagacherie P, Legros J.-P, Rounsevell M. 1999. Modelling wheat yield responses to soil and climate variability at the regional scale. *Climate Research*, 11(3): 209–220.
- Wei M, Sung A H, Cather M E. 2004. Detecting Spatial Outliers Using Bipartite Outlier Detection Methods.

- In: *IKE*: 236–244.
- Wong M, Asseng S. 2006. Determining the causes of spatial and temporal variability of wheat yields at sub-field scale using a new method of upscaling a crop model. *Plant and Soil*, 283(1): 203–215.
- Wu Z, Yao F, Zhang J, Liu H. 2024. Estimating Forest Aboveground Biomass Using a Combination of Geographical Random Forest and Empirical Bayesian Kriging Models. *Remote Sensing*, 16(11): 1859.
- Xie Y, He E, Jia X, Bao H, Zhou X, Ghosh R, Ravirathinam P. 2021. A statistically-guided deep network transformation and moderation framework for data with spatial heterogeneity. In: 2021 IEEE international conference on data mining (ICDM): 767–776.
- Xiong W, Conway D, Holman I, Lin E. 2008. Evaluation of CERES-Wheat simulation of Wheat Production in China. *Agronomy Journal*, 100(6): 1720–1728.
- Xu E, Wang R, Zhang H, Yu Z. 2019. Coupling index of water consumption and soil fertility correlated with winter wheat production in North China Region. *Ecological indicators*, 102: 154–165.
- Xu Q, Liang H, Wei Z, Zhang Y, Lu X, Li F, Wei N, Zhang S, Yuan H, Liu S, et al. 2024. Assessing climate change impacts on crop yields and exploring adaptation strategies in Northeast China. *Earth's Future*, 12(4): e2023EF004063.
- Yang W, Deng M, Li C, Huang J. 2020. Spatio-temporal patterns of the 2019-nCoV epidemic at the county level in Hubei Province, China. *International journal of environmental research and public health*, 17(7): 2563.
- Yin X, Zhang W, Jing X. 2023. Static-dynamic collaborative graph convolutional network with meta-learning for node-level traffic flow prediction. *Expert Systems with Applications*, 227: 120333.
- Yu H, Kim S. 2012. SVM Tutorial-Classification, Regression and Ranking. *Handbook of Natural computing*, 1: 479–506.
- Yujun C, Juhua P, Jiahong D, Yue W, Zhang X. 2019. Spatial-temporal traffic outlier detection by coupling road level of service. *IET Intelligent Transport Systems*, 13(6): 1016–1022.
- Zampieri M, Toreti A, Ceglar A, Naumann G, Turco M, Tebaldi C. 2020. Climate resilience of the top ten wheat producers in the Mediterranean and the Middle East. *Regional Environmental Change*, 20(2): 41.
- Zeng X, Li Z, Zeng F, Caputo F, Chin T. 2023. Spatiotemporal Evolution and Antecedents of Rice Production Efficiency: From a Geospatial Approach. *Systems*, 11(3): 131.
- Zhang C.-t, Yang Y. 2019. Improving the spatial prediction of soil Zn by converting outliers into soft data for BME method. *Stochastic Environmental Research and Risk Assessment*, 33: 855–864.
- Zhang H, Tang Y, Chandio A A, Sargani G R, Ankrah Twumasi M. 2022. Measuring the effects of climate change on wheat production: evidence from Northern China. *International Journal of Environmental Research and Public Health*, 19(19): 12341.
- Zhang J, Wang B, Hua M, Chen Z, Liang S, Kang X. 2024. Spatiotemporal Meteorological Prediction Based on Fully Convolutional Neural Network. *IEEE Transactions on Geoscience and Remote Sensing*,
- Zhang W, Xi M, Liu H, Zheng H. 2023. Low sensitivity of net primary productivity to climatic factors in

- three karst provinces in southwest China from 1981 to 2019. *Ecological Indicators*, 153: 110465.
- Zhang Y, Li B. 2022. Detection of the spatio-temporal differentiation patterns and influencing factors of wheat production in Huang-Huai-Hai region. *Foods*, 11(11): 1617.
- Zhang Y. 2012. Support vector machine classification algorithm and its application. In: International conference on information computing and applications: 179–186.
- Zhang Z, Li Z, Song Y. 2024. On ignoring the heterogeneity in spatial autocorrelation: consequences and solutions. *International Journal of Geographical Information Science*, 38(12): 2545–2571.
- Zhang Z, Song Y, Karunaratne L, Wu P. 2024. Robust interaction detector: a case of road life expectancy analysis. *Spatial Statistics*, 59: 100814.
- Zhang Z, Zhang F, Du J, Chen D, Zhang W. 2021. Impacts of land use at multiple buffer scales on seasonal water quality in a reticular river network area. *PLoS One*, 16(1): e0244606.
- Zhu A.-X, Lu G, Liu J, Qin C.-Z, Zhou C. 2018. Spatial prediction based on Third Law of Geography. *Annals of GIS*, 24(4): 225–240.
- Zhu H, Hao H.-K, Lu C. 2024. Enhanced support vector machine-based moving regression strategy for response prediction and reliability estimation of complex structure. *Aerospace Science and Technology*, 109634.
- Zhu K, Yang X, Zhang Y, Liang M, Wu J. 2024. A Heterogeneity-Aware Car-Following Model: Based on the XGBoost Method. *Algorithms*, 17(2): 68.

致 谢

时光荏苒，岁月如梭，转眼间，我的博士研究生生涯即将结束，回首这几年的求学之路，感慨万千。在我的研究生生涯中，我得到了许多人的帮助和支持，他们的帮助和支持是我走过这段求学之路的动力和支撑。在此，我要向所有关心、支持和帮助过我的人表示衷心的感谢。

首先，我要感谢我的导师于强研究员，他在我求学期间给予了我无微不至的关怀和指导。在我遇到困难和问题时，他总是耐心地给予我指导和帮助，帮助我分析我适合做哪个方向，适合做什么研究内容。在于老师的鼓励和支持下，我有机会获得国家留学基金委资助的两年澳大利亚访学机会。于老师经常为同学们分享他过去的科研和工作经历，以及人生中各个重要阶段的导师和朋友，于老师总是能洞悉当今最前沿和亟需解决的科学问题，为我的科学研究提供了宏观的指引。于老师严谨的治学态度和严格的工作作风，深深地影响了我，使我受益匪浅。感谢于老师在我求学期间给予我的指导和帮助，让我在科研的道路上不断前行。祝愿于老师未来的学术研究取得更大的成就，培养出更多优秀的学生，祝于老师身体健康，万事顺心。

同样，我也要感谢生态模拟课题组、资源环境学院和水保所各位老师，感谢岳超教授，吕肖良教授，赵刚教授，王斌教授，何建强教授，李毅教授，冯浩研究员，刘准桥教授，靳宁副教授，姚宁副教授，徐守华老师，张明希博士，朱清高子博士等，以及国家气象中心高级工程师何亮博士；感谢课题组的师兄师姐们：李林超，薛金儒，贾倩兰，戎亮秉，王亚凯，张艳，赵杰，何沁锶，梁楚涛，何杰，庄伟，刘亚东，史利洁，李思逸等；感谢同届的施宇，李聪聪，赵洪飞，刘锦月，杨晶晶，柏延文；感谢各位师弟师妹们：田琪，李鹏，王雪彤，谭姣姣，陈彬，康晓凤，闫雨杏，郭晨辉，姚林佳，王欢欢，侯鑫，朱琳等。感谢你们在我求学期间给予我的支持和帮助，感谢你们的友谊和陪伴，让我在科研的道路上不再孤单。课题组的每一位老师和同学都给了我很多的启发和帮助，让我在科研的道路上披荆斩棘。

尤其要非常感谢澳大利亚科廷大学的宋泳泽副教授，他在我的两年访学期间给予了我无私的指导和帮助。在这两年时间里，我见证了宋老师领导的科廷大学地理空间智能实验室（Geospatial Intelligence Lab）的建立，见证了实验室队伍一步步壮大，也一同作为组委会成员或工作人员参与举办了众多国际学术会议。宋老师的专业知识和严谨的学术态度深深地影响了我，使我在研究中受益匪浅。宋老师的悉心指导和耐心解答，从最基本的科研技能入手，包括论文撰写，科研绘图，代码编写和学术 Social 等技能的教授和提升，让我在这两年的国外学术研究中不断进步，取得了许多成果，参

加了很多国际会议，也认识了众多学术界和工业界的前辈／泰斗。宋老师不仅是我学术上的引路人，还在日常的工作生活中给予我无数关怀和照顾。他的为人处事也对我的人生观、价值观以及未来的规划有着深远的影响。在此，我要向宋泳泽老师表示最诚挚的感谢！祝您在未来的学术道路上取得更大的成就！

同时，我在澳大利亚两年的访学生活中，认识了许多来自世界各地的学者和研究人员，以及来自国内外不同大学，但有着共同目标和兴趣的朋友们。我要感谢澳大利亚科廷大学设计与建筑环境学院（School of Design and the Built Environment），和地球与行星科学学院（School of Earth and Planetary Sciences）的老师朋友和工作人员们，感谢他们在我的生活与学习中提供的帮助，感谢王翔宇教授、吴鹏教授、Joseph Awange 教授、Francesco Mancini 教授、许弘雷教授、**Mohammad Swapan** 副教授、Atiq Zaman 副教授、Ivana Ivanova 副教授，Vasilios Papastamoulis 博士、李生萍博士、Jeremy Wu 博士、孙浚博博士、Mr Zaid Osama Saeed、Miss Esra Çelik、Mr Imran Khan 等。感谢科廷大学作物与病害管理中心（Centre for Crop and Disease Management, CCDM）的 Mark Gibberd 教授，Julia Easton 博士，Arnold Salvacion 博士等。感谢西澳大学谢巧云博士及其团队。感谢莫道克大学任永林教授团队。

我还要感谢我国外课题组的朋友们，感谢王翔，陈春江，肖卓瑶，杨新月，张学渊，曹晓燕，李长春，曲瑞，王冰，王雨飞，商家婕，柴正媛，王杰龙，胡蛟，韦静娴，明雨佳，侯玮涛，谢明强，罗鹏，徐栋，郑鑫程，章鹏程，林港，陈来，谢慧黎，吴奕龙，茶明星，蒋林，赵卿颖，赵瑞华，Fawziah Almutairi, Ayesha Kaleem, Omri Yechezkel Ram, Tomáš Ič, Robert Andriambololonaharisoamalala 等同学和朋友们。感谢他们在学术交流和合作中给予我的支持和帮助，感谢他们的友谊和陪伴，让我在异国他乡感受到了家的温暖。

我最应该感谢的是生我养我的父母，他们一路的支持和鼓励是我不断前行的动力。感谢我的父母在我求学期间给予我的支持和帮助，让我在科研的道路上不断前行。感谢我的家人和朋友们，他们的关心和支持让我在求学期间感受到了温暖和力量。

前路依旧漫漫，但我将继续努力，不断前行。希望在未来的日子里，我能继续秉持严谨治学的态度，勇于探索未知领域，为科学研究贡献自己的力量。同时，我也希望能将我所学到的知识和经验传递给更多的人，帮助他们在科研的道路上勇攀高峰。

任凯

二〇二五年七月于 陕西 西安

个人简历

任凯，男，汉族，陕西商洛人。

2012年9月，考入西北农林科技大学资源环境学院地理信息系统专业，2016年6月本科毕业并获得理学学士学位。

2016年9月，进入南京师范大学地理科学学院（虚拟地理环境教育部重点实验室）地图学与地理信息系统专业攻读理学硕士学位，2019年6月顺利毕业。

2020年9月，进入西北农林科技大学资源环境学院土地资源与空间信息技术专业攻读农学博士学位，师从于强研究员。

2023年3月至2025年3月，作为访问研究助理（Visiting Research Associate）在澳大利亚科廷大学（Curtin University）地理空间智能实验室进行CSC博士生联合培养，合作导师为宋泳泽副教授。

发表的学术论文

SCI 期刊论文:

- [1] **Ren, K.**, Song, Y.*, Li, L., Mancini, F., Xiao, Z., Zhang, X., Qu, R., Yu, Q. *, 2025. Identifying climate and environmental determinants of spatial disparities in wheat production using a geospatial machine learning model. *GIScience & Remote Sensing*, 62(1): 2533487. <https://doi.org/10.1080/15481603.2025.2533487>. [中科院一区 Top, IF 6.9]
- [2] **Ren, K.**, Song, Y.*, Yu, Q. *, Second-dimension outliers for spatial prediction. *International Journal of Geographical Information Science*. [中科院一区 Top, IF 5.1] (Accept; in press)
- [3] Yang, X., Song, Y.*, Yoo C., **Ren, K.**, Wu, P., 2025. Irregular anisotropy in surface urban heat island footprint. *Sustainable Cities and Society*, p.106779. <https://doi.org/10.1016/j.scs.2025.106779>
- [4] Yang, L., Luo, P.*, Zhang, Z., Song, Y.*, **Ren, K.**, Zhang, C., Awange, J., Atkinson, P.M. and Meng, L., 2024. A spatio-temporal unmixing with heterogeneity model for the identification of remotely sensed MODIS aerosols: Exemplified by the case of Africa. *International Journal of Applied Earth Observation and Geoinformation*, 132, p.104068. <https://doi.org/10.1016/j.jag.2024.104068>
- [5] Zhang, X., Song, Y., Dewan, A., Guo, Z., Cao, X., Bie, Q., Xie, Y.*, Ma, X., **Ren, K.**, Zhang, H. and Xi, G., 2024. Ecological influence of oasisation on peripheral regions. *International Journal of Applied Earth Observation and Geoinformation*, 132, p.104004. <https://doi.org/10.1016/j.jag.2024.104004>

国际会议论文:

- [6] **Ren, K.**, Song, Y. and Yu, Q. (2025). A Novel Geospatial Machine Learning Approach for Uncovering Wheat Production Disparities: Evidence from Australia. *The 11th International Conference on Innovative Production and Construction (IPC2025)*. Perth, Australia.
- [7] **Ren, K.**, Song, Y. and Yu, Q. (2024). Assessing the Impact of ENSO and Indian Ocean Dipole on Australian Bushfires Using RandomForests. *2024 American Geophysical Union (AGU24)*. Washington, D.C.
- [8] **Ren, K.** & Song, Y. (2023). Local spatial heterogeneity of Wheat Yields in Western Australia. In *Proceedings of City+2023@Perth International Conference on Geospatial Big Data and Artificial Intelligence for Cities*. Perth, Curtin University.

学术交流

- [1] One of the **Organising Committees** of 11th International Conference on Innovative Production and Construction (*IPC2025, 4-9 August 2025, Perth, Australia*)
- [2] **International Conference Staff** of ISPRS Technical Commission IV Symposium 2024. (*October 22 - 25, 2024, Fremantle, Perth, Australia*)
- [3] One of the **Organising Committees** of 10th International Conference on Innovative Production and Construction (*IPC2024, 11-12 July 2024, Perth, Australia*)
- [4] One of the **Organizers and Committees** of an International Academic Conference, City+2023@Perth, Australia, with the theme of Geospatial Big Data and Artificial Intelligence for Cities. (*8-10 September 2023, Perth, Australia.*)
- [5] **Presentation** at the International Conference, City+2023@Perth, during Session 5 with the theme of Spatial methods. (*10 September 2023, Perth, Australia.*)
- [6] Attended an Australian-French Association for Research and Innovation (AFRAN) Forum at Murdoch University in Perth, showcasing and discussing current areas of cooperation as well as the potential for future enhanced collaboration. (*26 & 27 October 2023*)
- [7] Engaged in a Centre for Crop and Disease Management (CCDM) Precision Agriculture meeting with USyd PA Lab at Building 304, Curtin University. (*6 September 2023*)
- [8] Attended a workshop at Building 201, Curtin University, themed around Vicarious calibration of the NASA PACE mission: deployment of the MarONet ("Marine Optical Network") off Perth. (*20 October 2023*)

获奖情况

- [1] 参加第十一届“挑战杯”陕西省大学生创业计划竞赛，并获**省级铜奖**，组内成员不分先后（2022）
- [2] 参加西北农林科技大学第十届“挑战杯”大学生创业计划竞赛，并获**决赛银奖**，组内成员不分先后（2022）

* 本文全部基于 L^AT_EX 环境，使用 N_W^AFU Thesis 模板撰写与排版。