

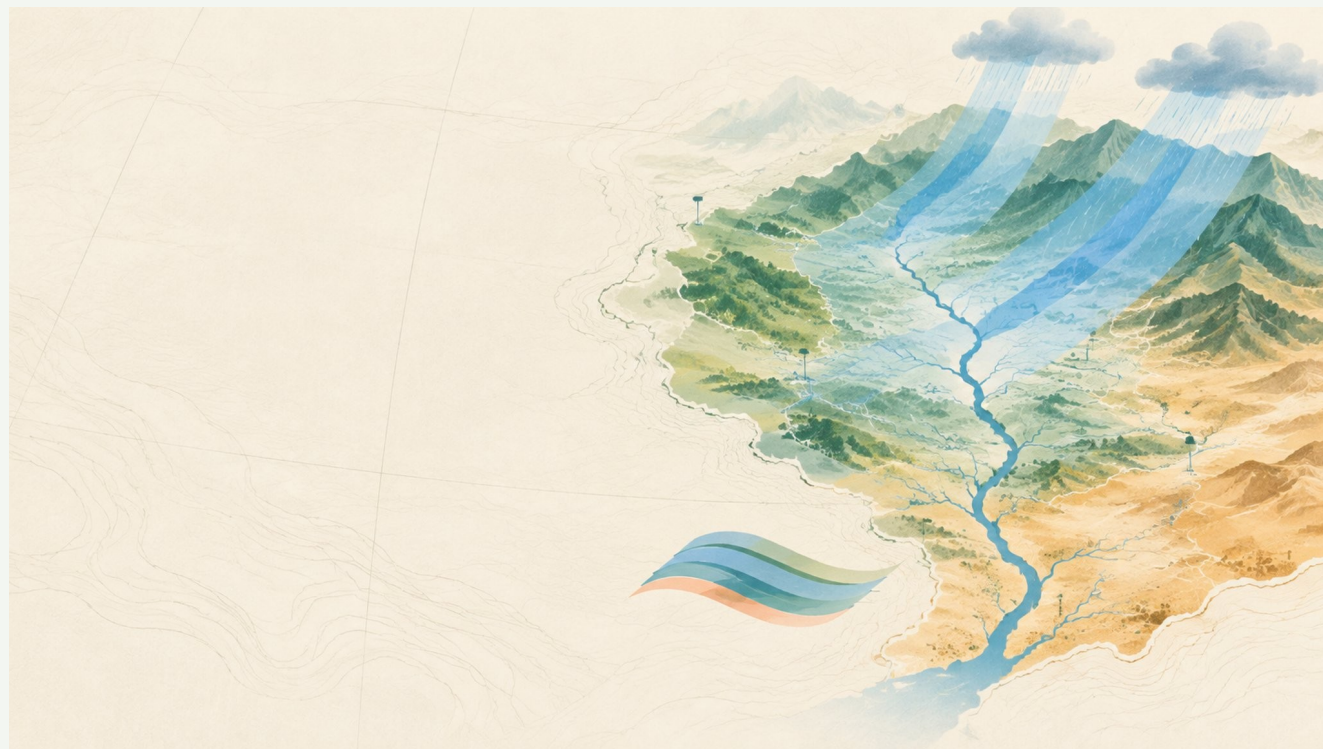
降水异常与 气象干旱聚集

SPI-3 × 干旱频率 × Moran / LISA

三个月累计降水如何变成
可比较的干旱异常指数？

Notebook 当前结果：固定种子的合成网格

CHIRPS 真实样片：只用于数据接入检查



今天要把三个问题说清楚

先问降水怎么变成指数，再问空间统计结果能不能叫“聚集”。

1

什么叫 SPI-3 ?

把最近 3 个月降水加起来，再按历史分布转成标准化异常。

证据：Notebook 的 SPI 网格图与 summary.csv

2

哪些地方更常达到干旱阈值?

本案例将 $SPI \leq -1$ 记为一次干旱月份，按有效月份求频率。

证据：频率图；阈值改变后的敏感性比较

3

象限颜色等于显著热点吗?

不等于。象限先描述高低组合，显著性要看置换 p 值和 FDR q 值。

本次合成示例的核对值

17.63%

平均干旱频率 ($SPI \leq -1$)

I = -0.0596

Global Moran 描述统计量

0 / 80

Local Moran 经 FDR 显著单元

从“降水少不少”变成可计算的问题

先固定空间单元和时间尺度，算法才知道每个数代表什么。

环境问题

哪里更常偏干？

相邻网格的偏干情况会一起出现吗？

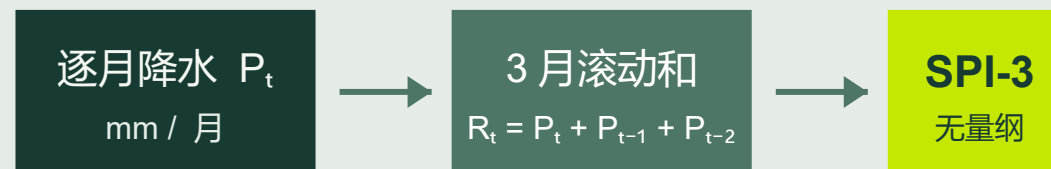
对应的统计对象

空间单元：网格像元

时间单位：月；SPI 时间窗：3 个月

课堂数组：10 年 × 12 月 × 80 格
网

从输入到可比较指标



然后统计：SPI ≤ -1 的月份占比 → 干旱频率图

再检查邻域关系：Global Moran I / Local Moran

课堂图件来自合成数组；真实 CHIRPS 只用于数据练习。

读到任何数值，都要能回答：哪个网格？哪几个月？用的是什么基期和阈值？

先分清三种 “数据”

产品、随包样片、Notebook 课堂输入用途不同，不能混写。

01 公开产品

CHIRPS v2.0

卫星估计结合站点资料的
降水格网产品。

月降水；常用单位 mm
全球数据体量大，按官方入口获取

入口：[data.chc.ucsb.edu/
products/CHIRPS-2.0/](http://data.chc.ucsb.edu/products/CHIRPS-2.0/)

02 随包真实样片

3 个月 × 1 个小区域

2023-11、2023-12、2024-01
全球 GeoTIFF 与裁剪样片

裁剪：110.5–111.5°E、27–28°N
20×20 像元；EPSG:4326

用于核查文件、空间范围和元数据
不够长，不能直接支撑本案 10 年 SPI

03 当前 Notebook 输入

合成教学序列

固定随机种子

10 年逐月 × 80 格

🌐 Notebook 默认不读 CHIRPS

所有结果图均标注 `synthetic_demo`

真实样片不是课堂结果的输入

下载得到数据 ≠ Notebook 已使用数据。报告中要写清当前运行的数据状态。

每一步都留下可检查的结果

按顺序跑完：知道输入是什么，也知道结果保存在哪里。



每道检查点都能找到证据

输入与 QC Notebook 输入摘要 来源、单位、时间、缺失	算法结果 SPI-3、干旱频率 Global / Local Moran	文件输出 outputs/student_run/figures/ summary.csv 与 LISA 诊断表	课堂结论 只描述本次合成运行 不外推真实干旱分布
---	---	---	---------------------------------------

出错时先定位是哪一道检查点失败，再决定要不要继续运行。

开始计算前，先过数据质量关

先确认每个数字代表什么；真实产品接入时，再补做产品级检查。

当前 Notebook：合成教学输入

1

形状对不对？

120 个月 × 80 个格网；输出网格为 8 行 × 10 列。

2

单位和状态写清了吗？

这是固定种子合成降水序列；不是 CHIRPS 实测值。

3

时间顺序和缺失值检查了吗？

确认月序列连续；真实数据要检查 NoData、掩膜和有效月数。

4

空间支持一致吗？

同一网格、同一行列方向；真实栅格另核 CRS 与分辨率。

判断：当前可继续演示算法

合成数组

Notebook 直接生成，随机种子 102。

每个格网有 120 个月的教学序列。

接入真实 CHIRPS 前

先准备足够长的连续月资料，
再核产品版本、CRS、单位、NoData、
时间范围和空间掩膜，并重新运行。

如果输入状态、时间轴或像元位置说不清，先停下来，不要先解释地图。

SPI-3：从降水到标准化异常

先把三个月降水合在一起，再与同一历月的基线比较。

STEP 1 滚动累计

$$R_t = P_t + P_{t-1} + P_{t-2}$$

每个像元单独计算。

前两个月没有完整 3 月窗，记为 NaN。

STEP 2 按历月拟合

同一个日历月单独建基线

正累计值拟合 Gamma；位置固定为 0。

零值概率 q_0 单独保留。

STEP 3 转成标准正态分数

$$H(x) = q_0 + (1 - q_0)G(x)$$

$$\text{SPI-3} = \Phi^{-1}(H(x))$$

SPI 无量纲；负值表示低于本历月基线。

代码里的回退条件不能省略

每历月有效样本不足

少于 $\text{min_samples} = 5$

不能稳定拟合 Gamma

正值样本不够或无变异

正值少于 2 个、方差近零

或拟合报错时回退到 z-score

解释时怎么写

结果表标明回退方法；

不可说它等同于完整 SPI 拟合。

本例只有 10 年合成序列，用来教计算链；真实应用要有足够长且质量合格的历史基线。

两类空间统计，回答两个不同问题

比较前先核对：输入变量不同，检验方式也不同。

全局：整体上是否呈现空间相似？

Global Moran I

变量：120 个月中 $SPI-3 \leq -1$ 的干旱频率

本次 $I = -0.0596$ ；是描述统计量

代码未做置换或显著性检验，不能写成“显著随机分布”。

Local 象限怎样读

- HH：高值周围也偏高
- LL：低值周围也偏低
- HL / LH：局部空间离群

局部：哪些像元的值与邻域组合通过检验？

Local Moran I + 置换检验 + BH-FDR

变量：末年 12 个月平均 $SPI-3$ ；置换 4,999 次，随机种子 102。

先看 HH / LL / HL / LH 象限，再只把 $q \leq 0.05$ 的像元称为显著。

当前合成运行：80 格网中 0 个通过 FDR；灰色代表本次未达显著阈值。

不要把象限当成显著性

象限只按本地值和邻域空间滞后的正负分类。显著性要看置换 p 值，再看多重检验校正后的 q 值。

两种变量、两套问题，不可互换。

本页所有数值均来自固定种子合成教学数据，不代表真实区域诊断。

按这个顺序运行 Notebook

从案例文件夹打开，不要先改数据或跳到结果图。

1

打开核心 Notebook

cases/02_chirps-drought-hotspots/notebook/ 核心算法代码 .ipynb

2

选择案例环境，从第一格运行

在 VS Code / Jupyter 选择本案 Python 环境； Restart Kernel → Run All。

3

确认输入摘要和算法输出

先核合成数据标记，再看 SPI、频率、Moran/LISA 表和四张直接显示的图

4

保存自己的运行结果并记录参数

写明 data_status、随机种子、阈值与 FDR 规则；保留运行日志。

默认从合成序列运行；CHIRPS 样片不会自动接入模型。

运行完成后检查这些文件

outputs/student_run/

figures/01_spatiotemporal_overview.png

figures/02_trend_significance.png

figures/
03_spatial_diagnostic.png

figures/
04_sensitivity.png

summary.csv

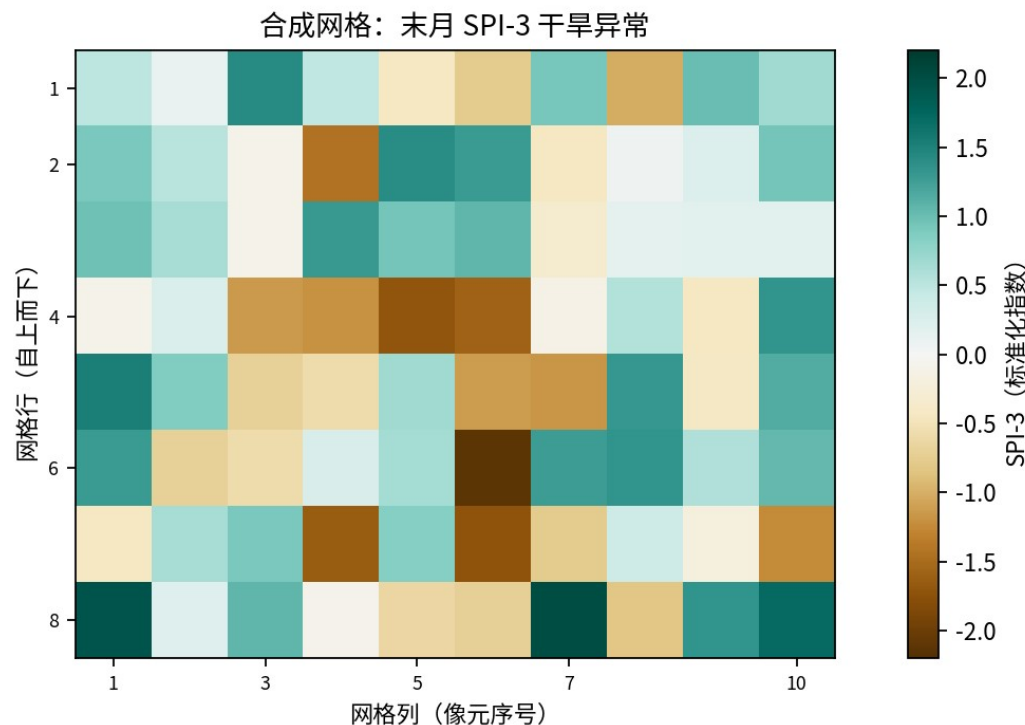
lisa_cell_diagnostics.c
sv

图、表都能打开后，再开始撰写解释。

逐格运行如遇导入或路径问题，先回到运行手册核对工作目录和环境。

结果一 | 末月 SPI-3 空间分布

先读标准化指数色带，再确认这个格网图表示什么。



怎么读

SPI-3 是无量纲异常

负值：低于该历月基线；

正值：高于该历月基线。

看空间位置

横纵轴是格网列、行编号，
不是经纬度坐标。

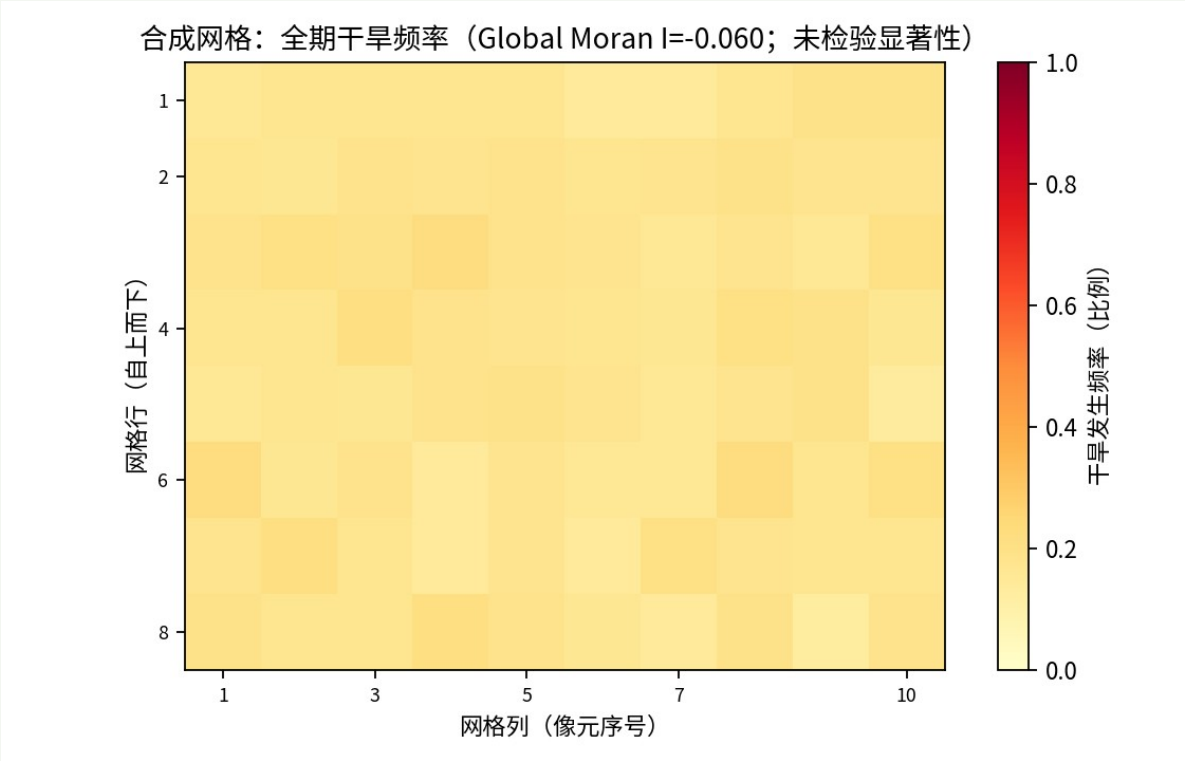
只显示末月结果

不能据此判断十年趋势。

数据状态：合成教学序列 | 8 × 10 格网 | 末月 SPI-3 | 不代表真实地区观测

结果二 | 全期干旱发生频率

先看阈值和统计时段，再看各格网的频率差别。



指标定义

干旱频率

SPI-3 ≤ -1 的有效月份数

本次合成输出的平均频率

17.63% 最高 22.03%

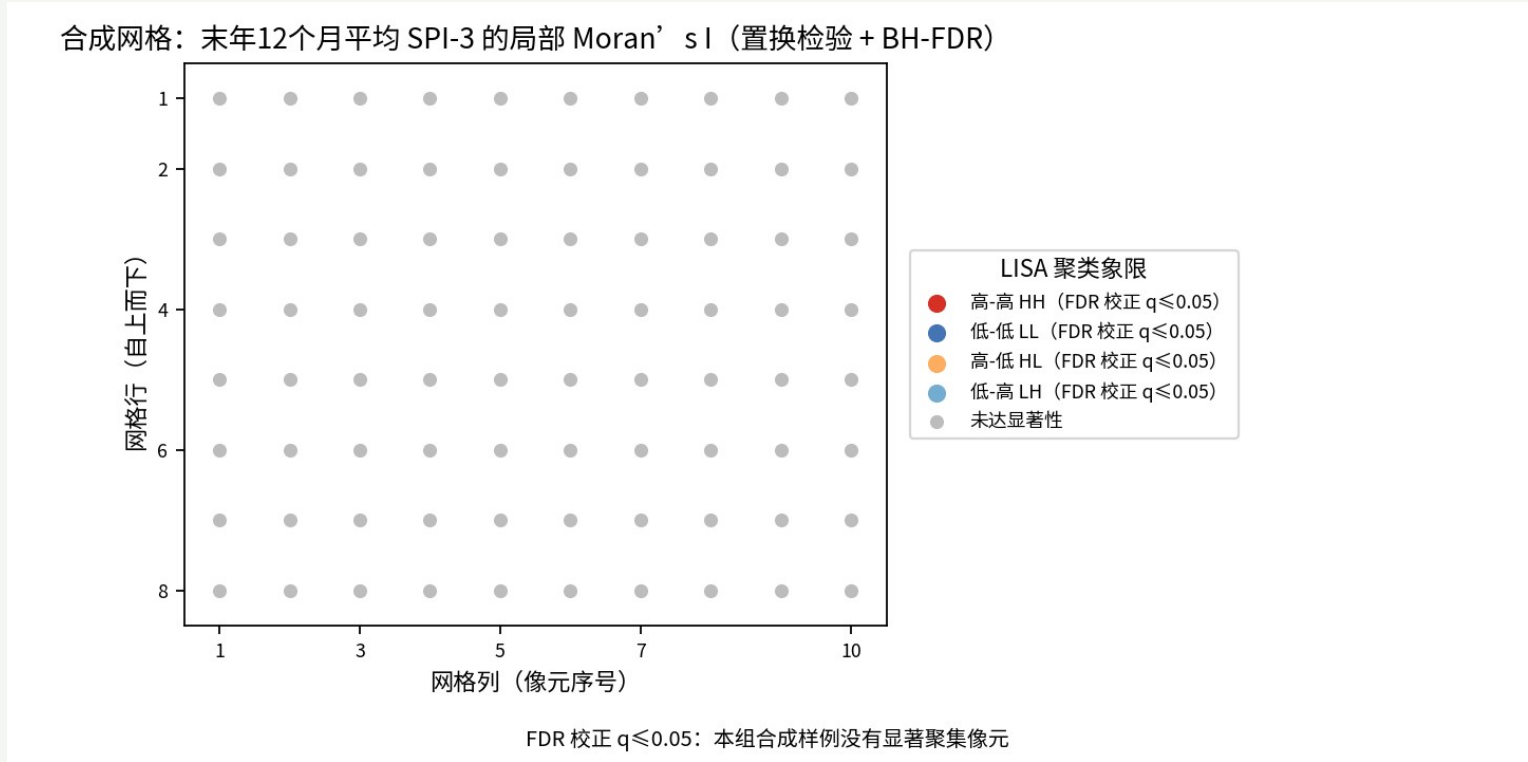
Global Moran I = -0.0596

仅为描述统计; Notebook 没有对它做显著性检验。

别把文件名中的 “ trend_significance” 当成图意：此图是频率地图，没有趋势显著性结果。

结果三 | Local Moran 象限与 FDR

象限先描述高低组合；只有通过多重检验校正，才标为显著。



本页分析对象

末年 12 个月平均 SPI-3

不是干旱频率图，也不是末月图。

置换检验：4,999 次

多重比较：BH-FDR

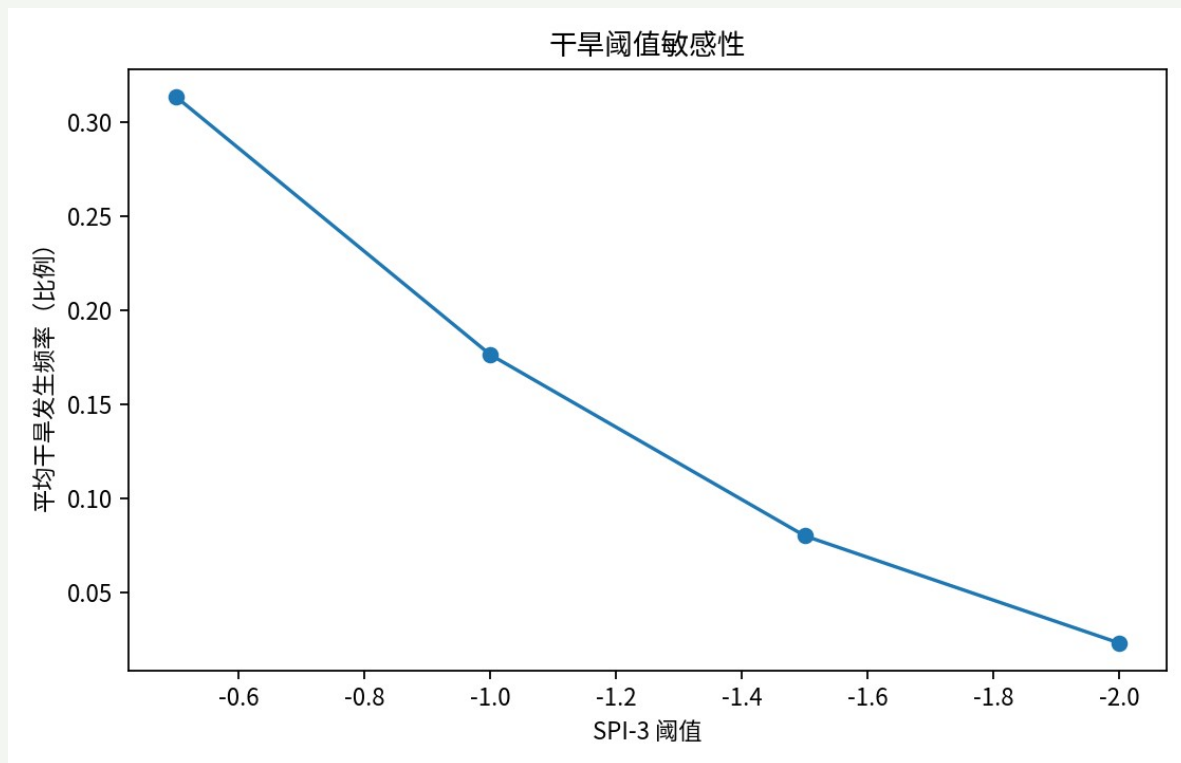
显著判据： $q \leq 0.05$

0 / 80 格网显著
灰点表示本次未通过 FDR 阈值。

正确说法：本次合成运行没有格网通过校正；这不等于真实研究区不存在局部聚集。

结果四 | 阈值改变，频率如何变化

阈值越严格，进入“干旱”统计的月份通常越少；这是一种敏感性检查。



比较的四个阈值

-0.5 - 1 - 1.5 - 2
同一份 SPI-3 输出，只改变判定线。

本次合成结果

-0.5 : 约 31.35%
-1.0 : 约 17.63%
-1.5 : 约 8.00%
-2.0 : 约 2.30%

频率随阈值变化，不是模型精度。

数据状态: synthetic_demo | 比较的是统计判定线 | 不能据此选择“最佳阈值”

可以这样解释，也要知道哪些话不能说

结果解释必须跟变量、时间窗口和数据状态一起出现。

可以支持的说法

“本次合成运行 中...”

本月各格网的 SPI-3 有高低差异；

全期 $SPI \leq -1$ 的频率空间分布如图；

Global Moran I = -0.0596，作为描述量；

Local Moran 的 80 个格网没有一个通过
BH-FDR $q \leq 0.05$ 。

每句都带“合成、变量、阈值或检验规则”。

不能直接支持的说法

“研究区发生 了...”

合成网格不是实测区域；

样片 CHIRPS 没有进入当前 Notebook；

没有真实观测对照、误差评估或因果设计；

不能把灰点写成“没有空间聚集”，
也不能把频率图写成趋势显著性图。

先写数据状态，再写结论。

如果一句话脱离了数据状态，回到 Notebook 和图件追溯表重新核对。

出问题时，按现象定位

先判断问题落在哪一步，再看对应文件；不要用改标题掩盖错误。

看到的现象	先核对什么	建议处理
Notebook 找不到 core_model	当前工作目录、案例根目录、Python 内核	从案例目录打开；重新选本案环境，然后 Restart Kernel 并从头运行。
图上全是灰点	lisa_cell_diagnostics.csv 的 p、q 值和显著列	若 q 都大于 0.05，这是当前阈值下的结果；不要手动涂出“热点”。
干旱频率与预期不符	SPI 阈值、有效月份、NaN 与 summary.csv	确认分母只含有效 SPI 时段，参数一致后再比较结果，不要把缺失期当成非干旱。
PPT 标题与图件不符	FIGURE_TRACEABILITY.csv 的 meaning 字段	先按追溯表改正含义，再检查 Notebook 文件名；文件名不能替代实际图意。

排查顺序：症状 → 证据文件 → 参数或环境 → 修复后重跑 → 记录差异。

交哪些结果，教师怎样验收

交的不只是图；每个结论都要能回到数据、参数和诊断表。

学生提交

01 四张结果图

末月 SPI-3、干旱频率、Local Moran、阈值敏感性

02 summary.csv

记录 data_status、均值、阈值、方法和校正参数

03 lisa_cell_diagnostics.csv

逐格提交 local I、p 值、q 值、象限和显著标记

04 简短结果说明

数据状态、参数、观察、限制；不把合成值写成实测结论

教师验收：逐项对照证据

数据状态有没有写对？

Notebook 默认是合成序列；CHIRPS 样片未自动接入。

算法变量和参数能否对上代码？

SPI-3、 $SPI \leq -1$ 、4999 次置换、BH-FDR $q \leq 0.05$ 。

图件、结论、追溯表是否一致？

频率图不写趋势显著性；象限不替代 FDR 判定。

限制是否明确？

按本案任务单评分量规验收，不另编评分权重。

目录建议：把所有学生结果放在 `outputs/student_run/`，并保留运行日志和参数记录。

用核心论文校准方法，不照搬结论

论文帮助我们理解 CHIRPS 产品背景；本案仍要以自己的数据、算法和追溯表为准。

论文

Funk et al. (2015)

介绍 CHIRPS 产品
与
极端降水监测背景。

DOI: 10.1038/sdata.2015.66

CC BY 4.0；随包保存公开全文。



本案实现

SPI-3 + Moran / LISA

固定种子合成序列，
用代码演示计算链和 QC。

不是论文研究区复现；
不是论文数值复现。



带走的问题

**如果换成真实资料，
还要补什么？**

连续历史基线、
产品 QA / NoData、
CRS 与空间尺度、
误差和敏感性评估。

课后任务： 在表格中列出论文与本案在“数据来源、空间尺度、SPI 基线、统计检验、推论边界”上的五个不同。

公开全文：<https://www.nature.com/articles/sdata201566> | 本地：论文资料 / 开放全文 / P02_Funk_2015_CHIRPS_CC_BY_4.0.pdf