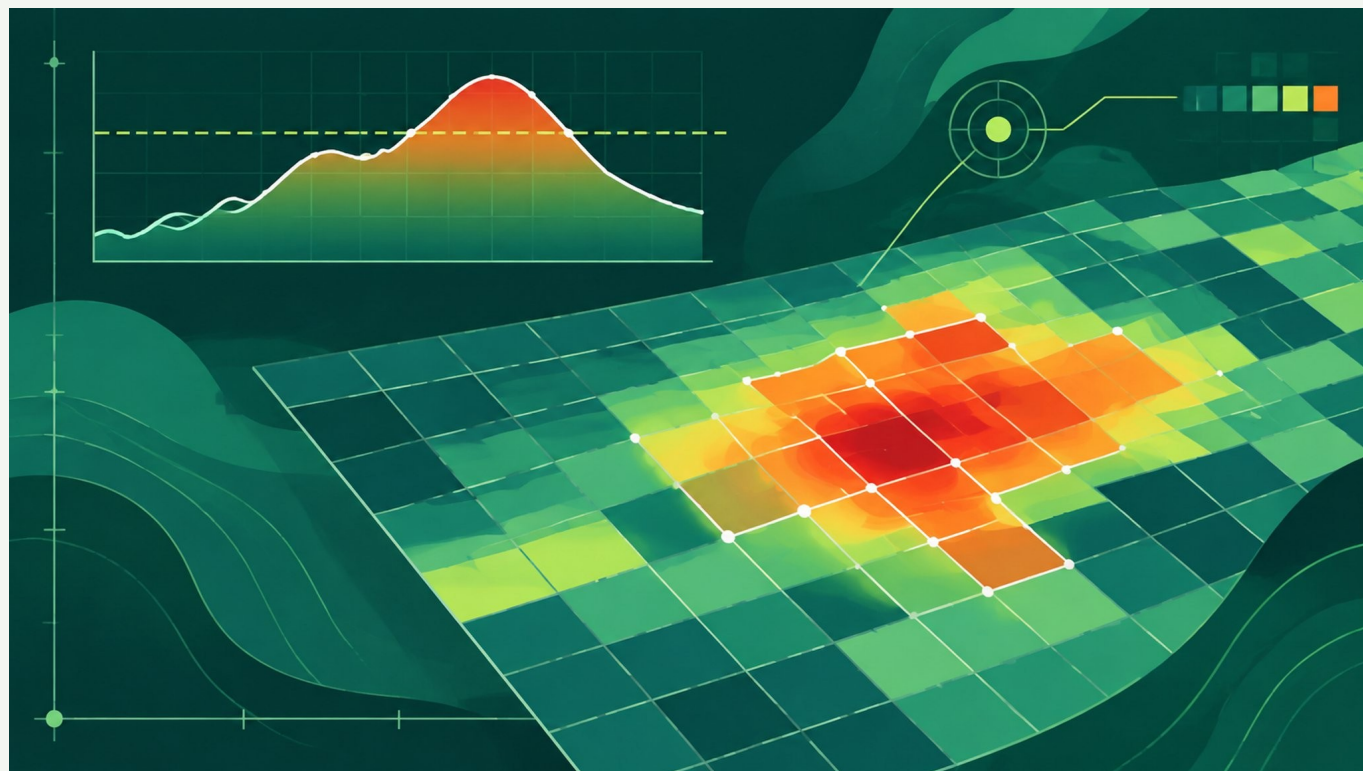


高温极端与热浪空间识别

百分位阈值 × 连续事件 × 空间汇总

超过多少度才算异常？
连续几天才算一次热浪？

Notebook 当前结果：固定种子的合成日最高温
ERA5-Land：只提供真实资料接入入口



今天要把三个问题说清楚

热浪识别不是只看“高温”，还要看基线、连续天数和统计窗口。

1

阈值从哪里来？

用每个网格前 365 日的最高温，算第 90 百分位阈值。

证据：单网格温度曲线、q90 虚线和 Notebook 参数

2

怎样才算一次热浪？

超过本网格阈值，并且连续至少 3 天，整段才会被标记。

证据：末年热浪日数和最长连续过程图

3

换阈值后结果会怎样？

比较 q85、q90、q95，但不把敏感性图写成趋势检验。

本案的固定口径

730 日 × 80 格网

前 365 日建阈值，后 365 日评估

q90 + ≥3 天

当前教学热浪定义

34.90°C

合成网格阈值均值

从“天气很热”变成可计算的问题

先确定网格、时间窗和阈值，再谈热浪日数。

环境问题

哪里更容易出现热浪？

一次高温和一段连续高温，统计含义不同。

统计对象

空间单元：80 个网格像元

时间单位：日；基线 365 日 + 评估 365

变量：日最高温，单位 $^{\circ}\text{C}$

从输入到热浪指标



事件 → 末年热浪日数、最长连续过程

再比较 q85 / q90 / q95 的敏感性。

课堂图件来自合成数组，不是 ERA5-Land 实测地图。

读到一个热浪数字，要能回答：哪一个网格？哪一段日期？阈值和持续天数是多少？

先分清三种 “输入”

产品资料、真实接入和课堂合成数组的用途不能混在一起。

01 公开产品

ERA5-Land

陆面再分析资料，常用
2m 温度等变量。

逐小时温度原始单位 K
需从 CDS 登录检索和下载

[cds.climate.copernicus.eu/
datasets/reanalysis-era5-land](https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land)

02 真实接入步骤

逐小时 → 日最高温

按 UTC 日界聚合逐小时 t2m
K 转 °C : $T - 273.15$

检查完整日、缺小时和时间连续性
保留请求参数、版本和 SHA-256

当前核心代码尚未实现文件接入

03 当前 Notebook 输入

合成日最高温

固定随机种子

730 日 × 80 格

网
前 365 日建阔

值
后 365 日统计热浪

所有结果图均为 synthetic_demo

下载真实资料 ≠ Notebook 已使用真实资料。报告中必须写清数据状态。

每一步都留下可检查的结果

输入、阈值、事件和图件要一一对应。



每道检查点都能找到证据

输入摘要 → 阈值数组 → 事件 mask → 热浪日数与最长过程 → 敏感性图 → 结果说明

目标目录: [cases/03_era5-heat-extremes/outputs/student_run/](cases/03_era5-heat-extremes/outputs/student_run/)

开始计算前，先过数据质量关

热浪识别最怕时间断档、单位混乱和基线偷看评估期。

当前 Notebook：合成教学输入

1

天数和网格数对吗？

730 个日序、80 个像元，输出排列为 8 行 × 10 列。

2

单位说清了吗？

合成值为 °C；ERA5-Land t2m 原始单位 K，接入前需换算。

3

时间轴连续吗？

真实逐小时资料需按 UTC 聚合日最高温，并检查缺小时。

4

基线与评估期分开了吗？

先用前 365 日求阈值，再只在后 365 日统计事件。

当前可以继续做算法演示

合成序列已固定

随机种子和输入由 Notebook 记录。

真实 ERA5 文件没有进入当前计算。

真实资料接入前

先聚合逐小时 Tmax、统一单位，
再检查网格、日序缺口和有效日数；
不要直接把月均温当成日最高温。

缺失日期、单位或基线范围没查清，就先停下来，不要先解释热浪地图。

算法一：每个网格先建立自己的 q90

同一条温度线，不同网格的高温基线可以不同。

逐格计算

$$T_{90,j} = \text{percentile}_{90}$$

j 表示一个网格；只用前 365 日
建立基线，不能把评估期混进阈值。

缺失值按 nanpercentile 忽略，
但仍要记录有效日数。

为什么不用一个全局温度阈值？

- 1 **不同位置的气候基线不同**
逐格 q90 表示“相对本地偏热”。
- 2 **百分位不等于固定摄氏度**
q90 是分布位置，数值会随网格变化。
- 3 **先保存阈值再识别事件**
结果表中应有每格阈值和有效日数。

检查点：阈值数组的长度必须等于网格数 80，不能只输出一个全局数。

算法二：把“热”变成连续事件

单独一天超过阈值还不够，连续规则决定哪些天真正计入热浪。

判断逻辑

`raw = T_eval >`

对每个网格分别扫描整段连续序列，短于 3 天的片段全部清零。

`event = 连续 True 的长度 ≥ 3`

一段序列示意

0 1 1 0 1 1 1

1 0
前两个 1 只有 2 天 → 不计

入
后四个 1 有 4 天 → 四天全部计
入

三个交付指标

热浪日数

评估期中 event=True 的天数

最长过程

每个网格的最长连续 True 长度

敏感性

q85、q90、q95 的热浪日占比

注意：当前代码统计的是后 365 日评估期；全 730 日 mask 只用于生成教学图和检查逻辑。

Notebook：按这个顺序运行和留痕

每一步都对应一个可检查的输出，遇到问题时先定位是哪一步。

1

加载依赖与字体

确认 NumPy、Pandas、Matplotlib 和中文字体可用。

2

生成或读取日最高温

当前 Notebook 固定种子生成 730×80 合成矩阵。

3

计算阈值与连续事件

先 percentile_threshold，再 heatwave_mask 和 longest_run。

4

保存图片件和摘要表

四张 PNG、summary.csv 和 FIGURE_TRACEABILITY.csv。

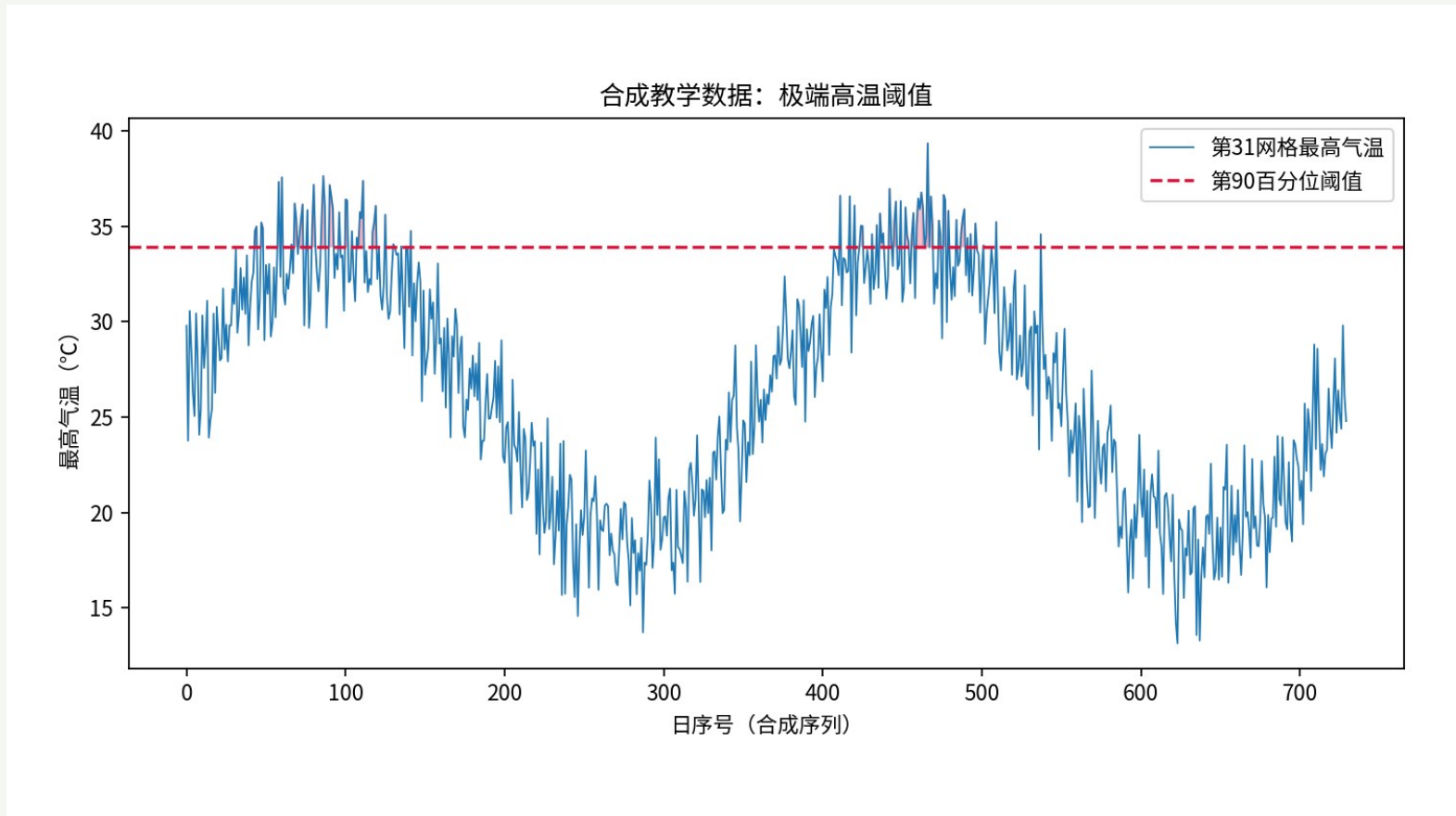
运行后的目录

```
notebook/figures/  
01_spatiotemporal_overview.png  
02_trend_significance.png  
03_spatial_diagnostic.png  
04_sensitivity.png
```

```
outputs/student_run/  
summary.csv + figures/
```

结果一：单格温度线与 q90 阈值

红色虚线来自前 365 日基线；阴影表示超过阈值的日子。



怎么读这张图

黑线 一个合成网格的日最高温

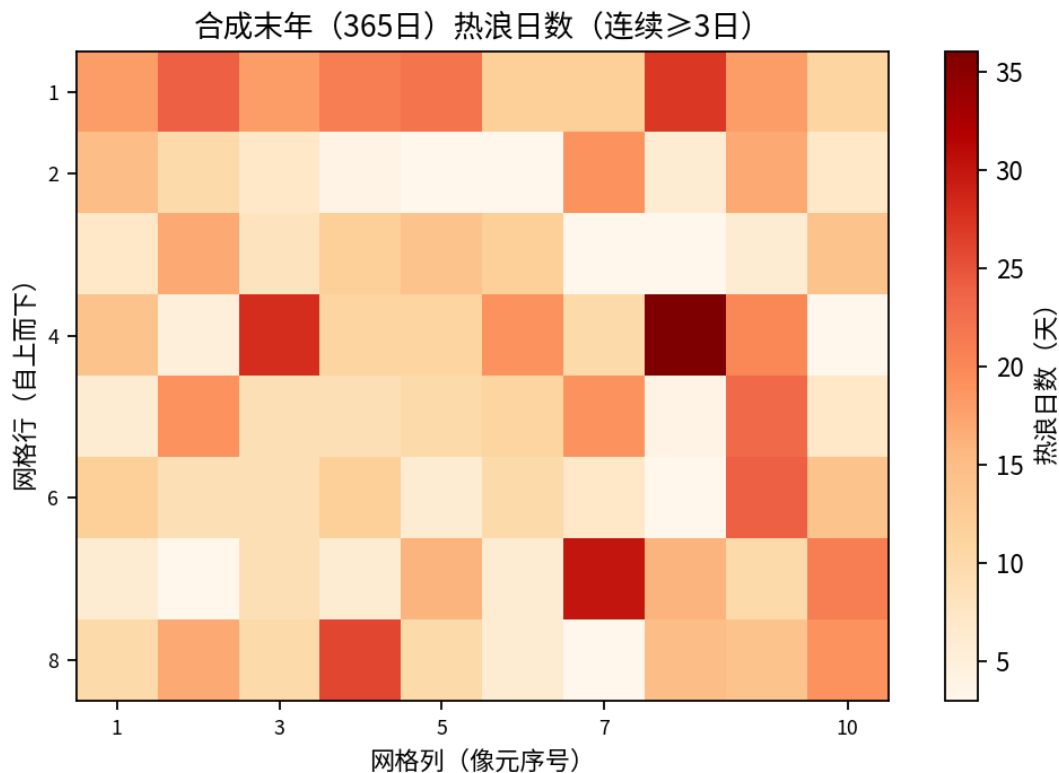
虚线 该网格前 365 日
q90

阴影 超过阈值的可疑高温日

这张图用于理解阈值，
不能单独证明热浪事件。

结果二：第二年哪些格网热浪日更多？

颜色表示每个合成格网在 365 日评估期内满足规则的热浪天数。



读图时看三件事

格网 8×10 教学数组，没有真实地理坐标

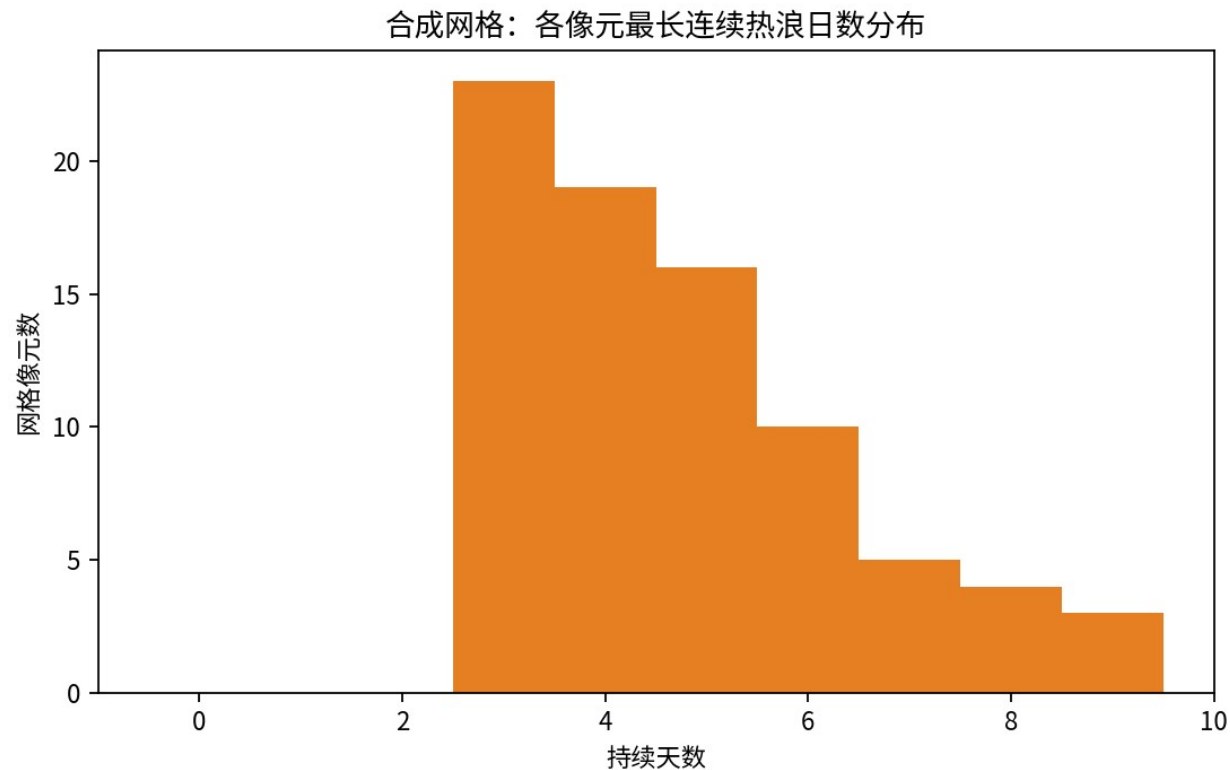
颜色 颜色越深，合格热浪日越多

时间 仅统计后 365 日评估期

这不是某地区地图，
不应据此作真实地点结论。

结果三：最长连续热浪过程有多长？

这页是 80 个格网的持续时间直方图，不是空间分布地图。



读图要点

横轴 最长连续热浪日数

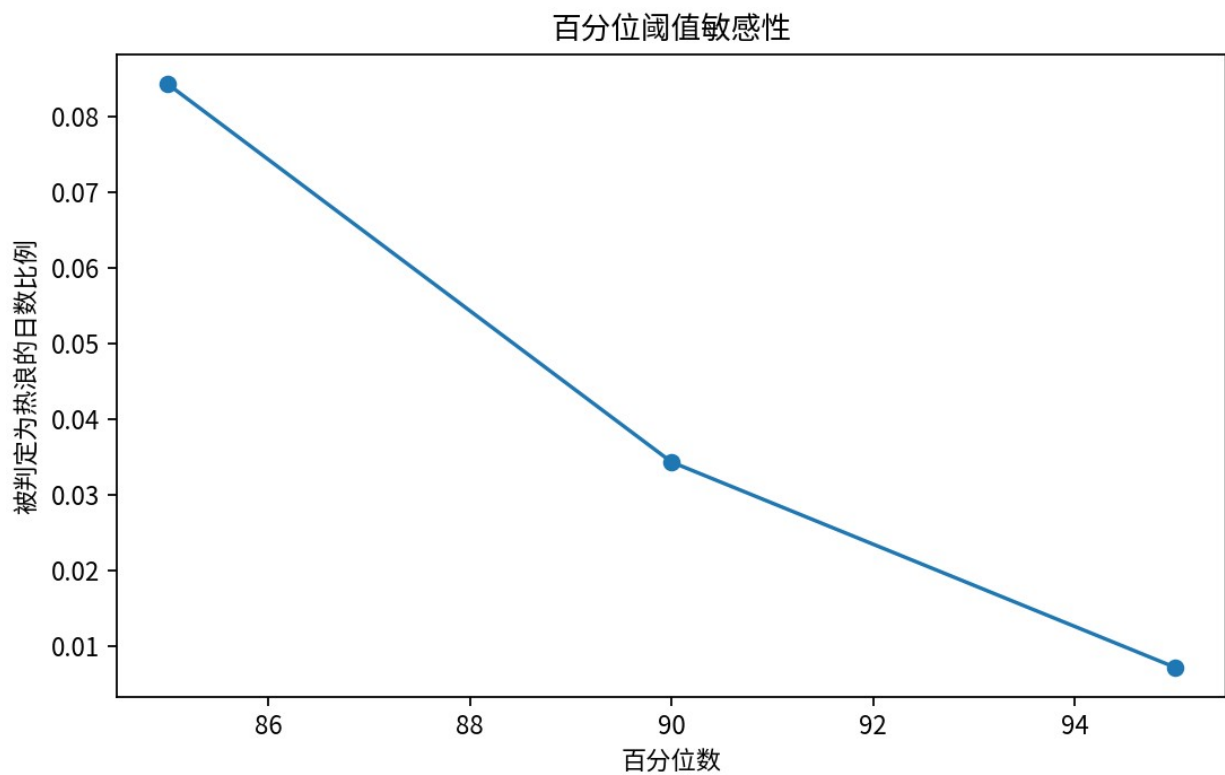
纵轴 落在该时长的格网数量

当前摘要 最大连续过程为 9 天

想看空间差异，使用上一页的格网图；不要把两种图混称。

结果四： 阈值改变， 热浪比例怎样变？

同一评估期下比较 q85、 q90、 q95， 观察阈值敏感性。



当前合成结果

q85 约 8.43% 的评估日

q90 约 3.43% 的评估日

q95 约 0.72% 的评估日

敏感性回答“参数变了会怎样”，
不是趋势检验，也不是显著性检验。

哪些结论可以说，哪些不能说？

把算法验证结果和真实生态环境发现分开，是本案最重要的科学边界。

当前 Notebook 可以支持

- ✓ 验证逐格百分位阈值的计算顺序
- ✓ 验证连续不少于 3 天的事件筛选
- ✓ 比较评估期热浪日数和最长过程
- ✓ 观察 q85、q90、q95 的敏感性

数据状态：synthetic_demo；固定随机种子 103。

当前 Notebook 不能支持

- × 某地区真实热浪空间格局
- × 多年热浪趋势或显著性检验
- × 健康暴露、风险或因果结论
- × 直接替代 ERA5-Land 数据处理

接入真实资料后，必须重新做 QC、运行和解释。

常见故障：先定位步骤，再改参数

出现异常时按现象排查，不要为了“看起来合理”直接改结果。

现象	先检查	解决办法
模块导入失败	是否从案例根目录打开 Notebook ？	选择本案例环境，从第 1 格重启并运行。
图中文显示方框	Matplotlib 是否找到中文字体？	安装可用字体，重启内核后重新运行。
结果和本次摘要不同	随机种子、代码版本和输入是否一致？	固定 seed=103 ，记录改动后再比较。
全图没有热浪	检查阈值方向是否为严格大于和连续 ≥ 3 日。	不要删掉持续性筛选来“造出”事件。

真实 ERA5-Land 若缺日期或单位不是 $^{\circ}\text{C}$ ，先修正输入；当前 Notebook 不会替你自动清洗真实文件。

提交什么，老师如何验收？

交付物要能追溯：别人拿到你的文件，能看懂你做了什么、用的是什么口径。

学生提交清单

- 01 运行后的 Notebook
- 02 四张直接显示的 PNG 图
- 03 outputs/student_run/summary.csv
- 04 参数、随机种子和运行截图
- 05 200–300 字结果解释

若接入真实数据，再附产品版本、下载日期、许可和数据处理记录。

验收指标

算法正确

q90、严格大于、连续 ≥ 3 天

图件完整

4 图齐全，标题与含义一致

结果可追溯

summary 与图件同一运行

边界说清楚

不把合成数组写成真实地图

复现可行

固定 seed，路径不依赖个人电脑

合成数据的课堂成绩只评价“算法和表达是否正确”，不能冒充真实区域研究成果。

论文桥接：把公开产品资料用在正确位置

本案借助论文理解 ERA5-Land 数据产品，不把论文结果说成本案的运行结果。

核心公开全文论文

ERA5-Land: a state-of-the-art global reanalysis dataset for land applications

Muñoz-Sabater et al. · 2021 · Earth System Science Data

13: 4349–4383 · DOI: 10.5194/essd-13-4349-2021

全文 HTML: essd.copernicus.org/articles/13/4349/2021/

PDF: essd.copernicus.org/articles/13/4349/2021/essd-13-4349-2021.pdf

开放许可：CC BY 4.0（引用与再使用请遵循原许可条款）

阅读时带着这三问

它讲什么？

ERA5-Land 的数据和评估。

本案用在哪？

核实产品背景、变量和来源。

哪些不能照搬？

论文数据结果不是本案合成结果；
本案阈值与 3 日规则是教学设定。