CNGF 5020 Mini Project

By **No-Burn Buddies: Straw Edition**

Collaborators:

Wen Lu

Xiaoyu Fang

Run Wang

Zijun Xu (Jimmy)

Mingyuan Chi

Presenters:

Mingyuan Chi

Zijun Xu (Jimmy)

Data Preprocessing

MODIS data cannot be directly used as it provides **not "straw burning points" but "fire points"**,

The original modis_ *. csv file contains all types of fires: forest fires, grassland fires, industrial fire points and even possibly volcanoes.

So we need to design a **classifier** to select the most likely fire points from these 1.07 million "fire points" to be straw burning.

Three rules for this classifier:

Data Preprocessing

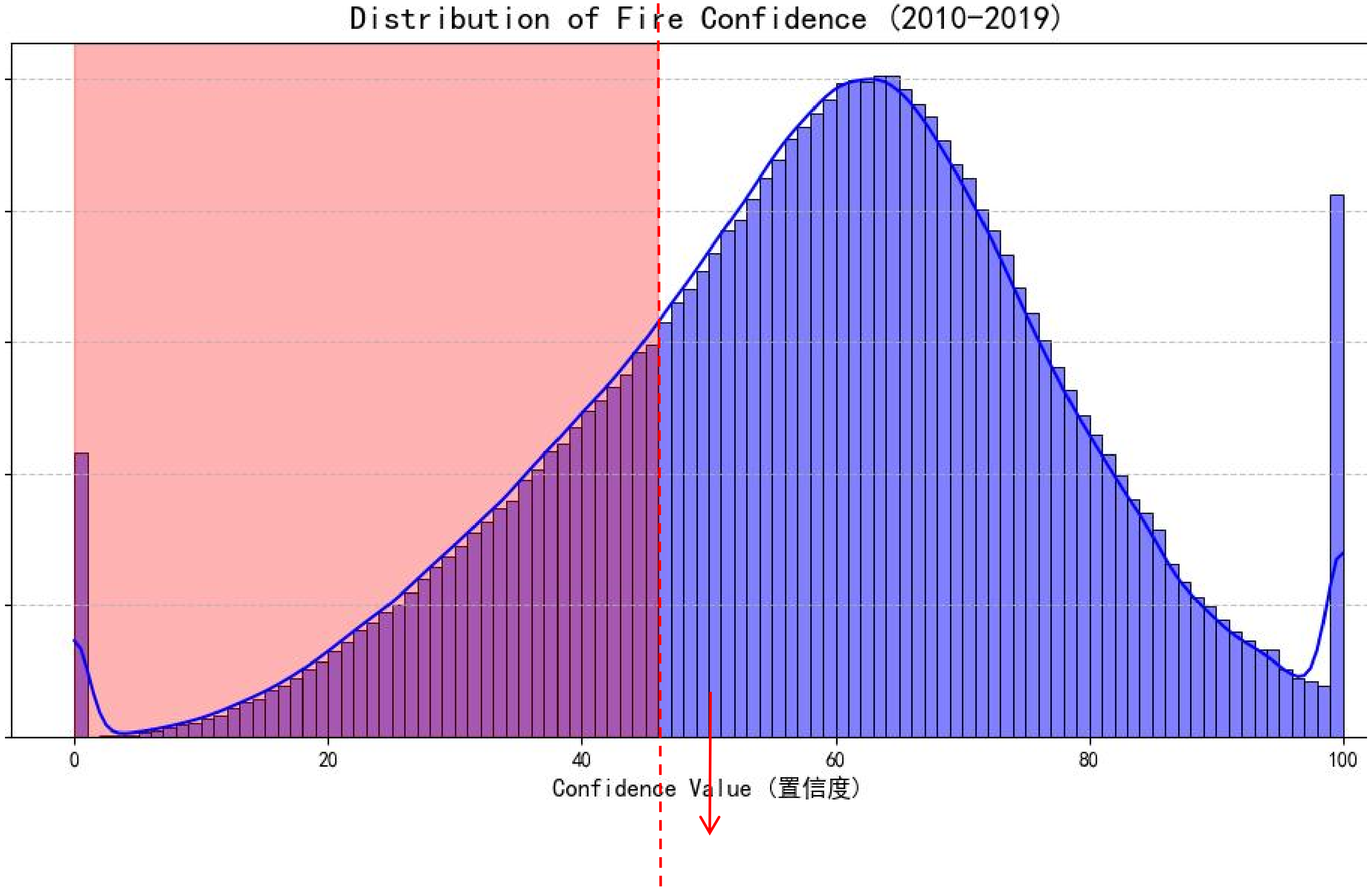
Rule 1: Data Quality Filter (Cleaning)

One round of cleaning, processing acq_date and acq_time, and deleting data with incorrect time formats.

Further cleaning, use **the confidence field**. The confidence in MODIS data indicates **the credibility of a real fire point**.

Suggested rule: We can filter out low confidence hotspots, for example, **only retaining hotspots with confidence>=50%**. This can filter out a large amount of "false fire" or noise.

Data Preprocessing



25% = 46.0: It means that out of 1.07 million fire points, there are a full 25% of them, with a credibility lower than 46%.
50% = 59.0: The median is 59%. Half of the data is below it, and half is above it.

Why 50?
This threshold (>=50) will help us filter out data with the lowest quality that exceeds 25% (because the 25th percentile is 46).
This accounts for at least 50% of the total data (as the median of 50% is 59).

It represents a clear stance – 'We only trust those hotspots that satellite teams consider to be "**moderately credible or above**".

Data Preprocessing

Rule 2: Space filter (where to burn)

Straw burning must occur on **farmland**.

When we receive the ignition point (e.g. longitude 125.5, latitude 46.5), we will search for the .tif files.

If this coordinate has no data in the .tif file (i.e. NoData), we reject it because it is not within the corn or wheat planting area.

Data Preprocessing

Rule 3: Time Filter (When does it burn)

Straw burning must occur **before or after the harvest period**.

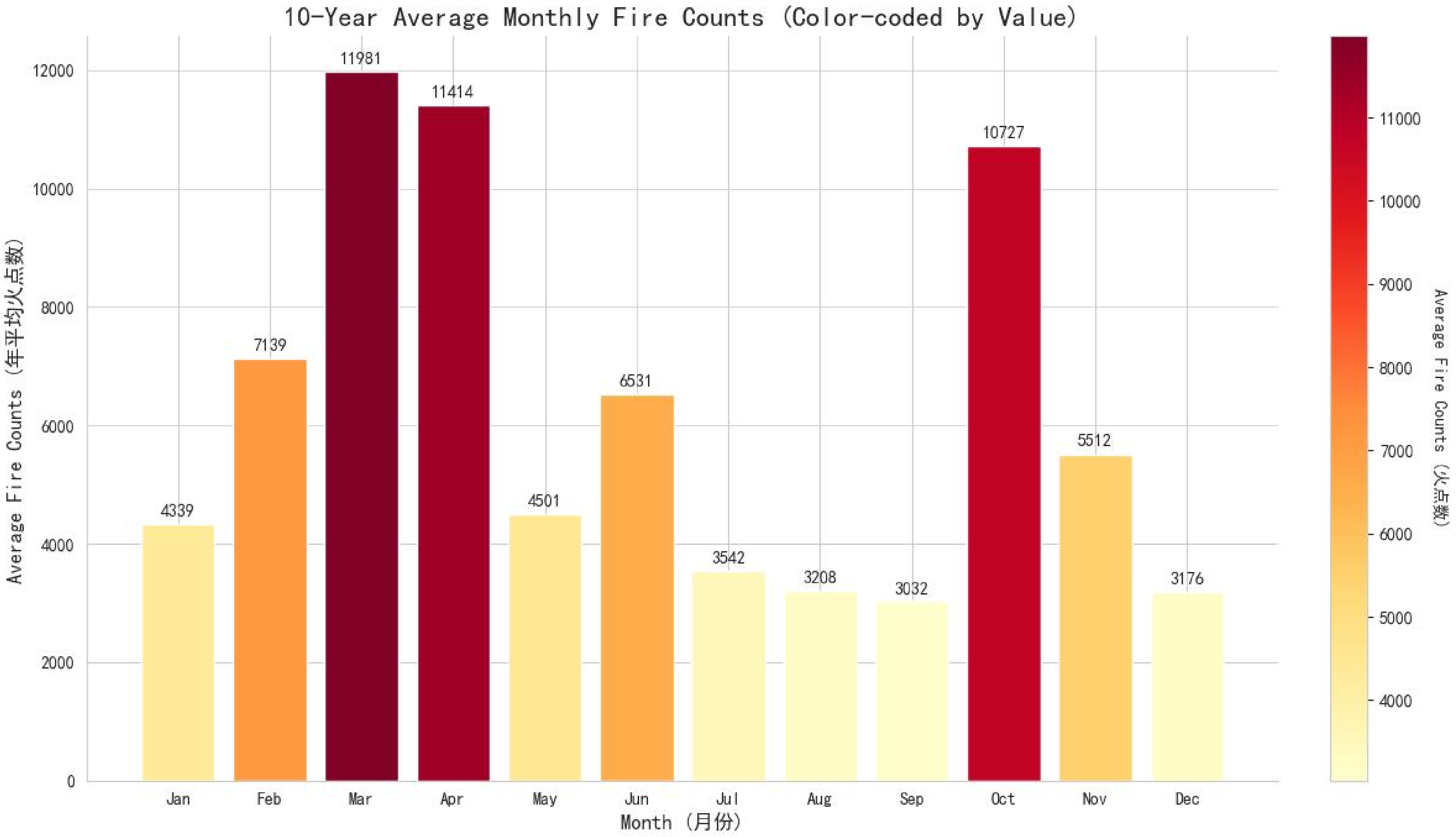
For fire points that have passed "Rule 2", we will extract the **"Crop Maturity Date (DOY)" of that pixel from the .tif file.

Is the fire near the mature stage? We use an appropriate time window, such as '15 days before harvest and 30 days after harvest'.

Data Preprocessing

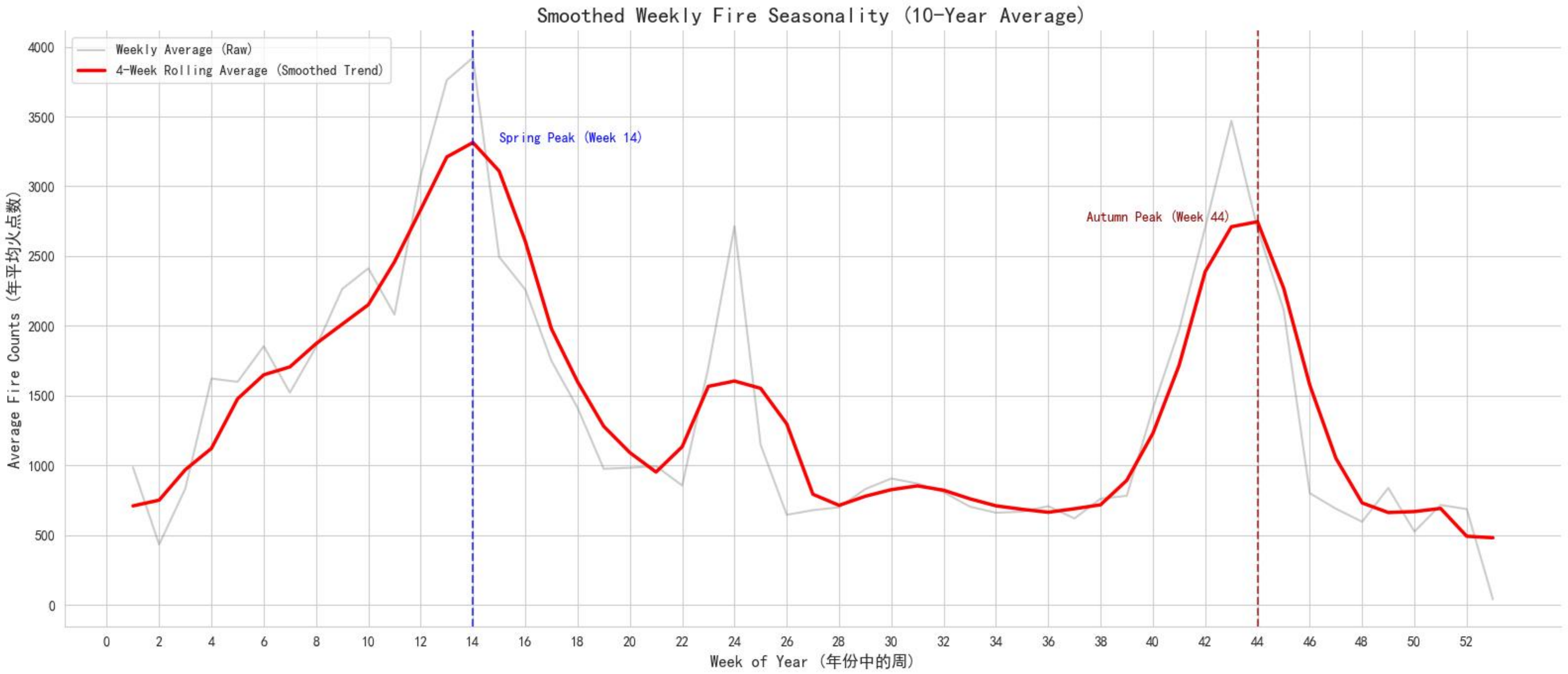
```
--- 目标: 应用 'confidence' 过滤器 (规则1) 清洗数据... ---
--- 清洗规则: 将只保留 confidence >= 50 的火点 ---
警告: 无法检测 __file__。假定文件在当前工作目录: /Users/mingyuan
正在加载 modis_2010_2019_processed.csv...
过滤前总火点数: 1073119
过滤后剩余火点数: 751014
成功移除了 322105 条低置信度火点 (30.02%)。

成功! 已将【已清洗】的数据保存到: /Users/mingyuan/modis_2010_2019_cleaned.csv
```

Task 1: Seasonality & Geographic Distribution



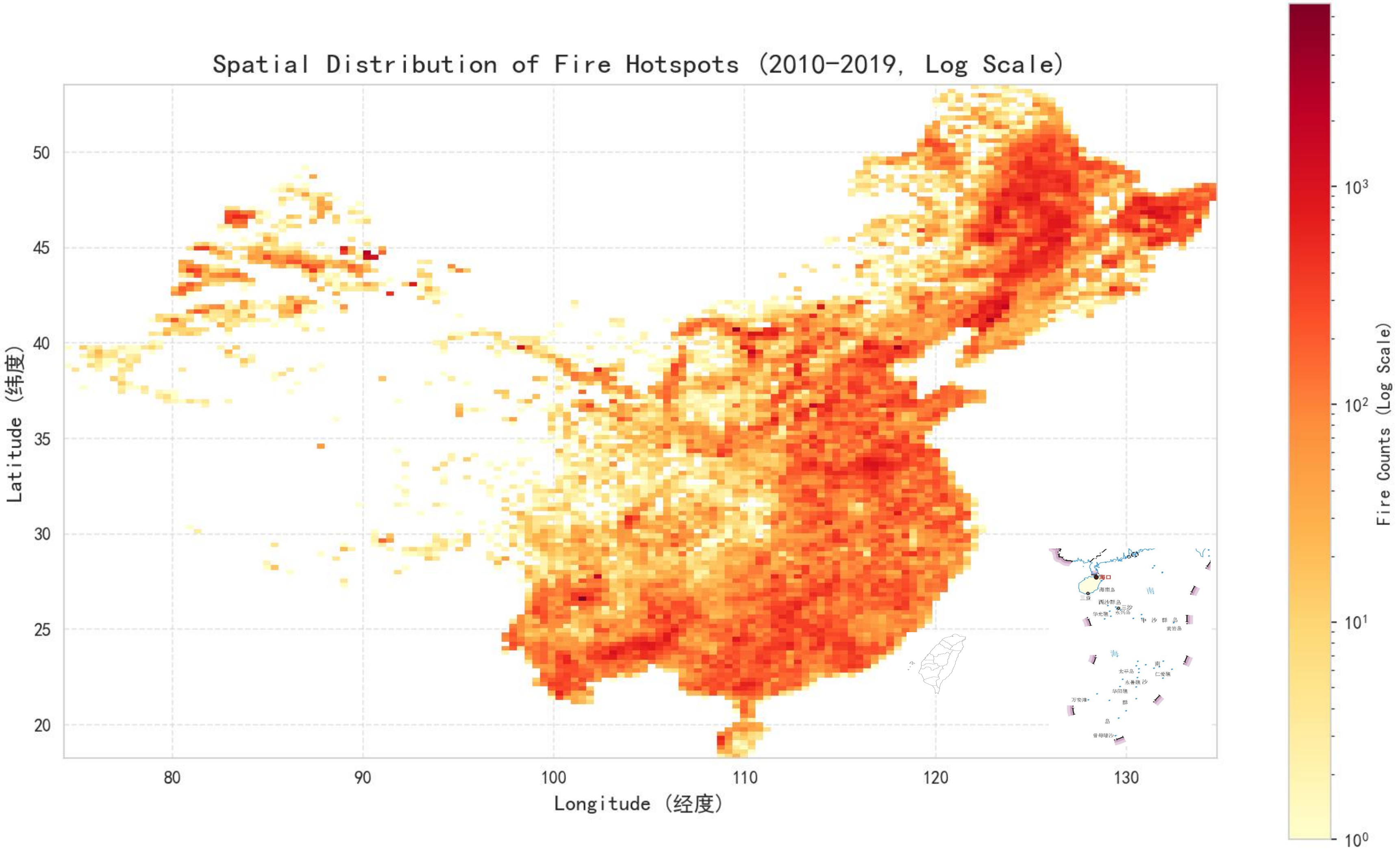
–Spring: **land preparation activities before spring planting**.
–Autumn: **post-harvest activities**.

Task 1: Seasonality & Geographic Distribution



–Why 53 weeks?
–Fires are not uniformly distributed across several months but are **concentrated in specific "hotspot weeks"**.

Task 1: Seasonality & Geographic Distribution



Conclusion: The spatial patterns are consistent with the temporal analysis. As Northeast China is t**he most important commercial grain base** (corn, rice), the high spatial concentration of fires further corroborates the hypothesis that these fires originate primarily from agricultural activities.

Task 2: What's the Percentage?

We addressed this question using the following steps:

1.Data Filtering: We **selected 135,512 high-confidence (confidence ⩾ 50) fire hotspots located within Heilongjiang Province** from the national MODIS fire dataset (2010-2019) .

2.Data Preparation: We **loaded the county boundary map for Heilongjiang (CHN_County.shp)** and the annual maturity date raster data (.tif) for maize and wheat from 2010 to 2019 .

3.Classification Rules: We applied the following rules to each fire hotspot:

- Rule 2 (Spatial): Did the fire occur within **a pixel identified as maize or wheat cropland** in the corresponding year's .tif file?

- Rule 3 (Temporal): If on cropland, did the fire's day of year (doy) fall within **the time window of [Maturity DOY - 15 days, Maturity DOY + 30 days]** for that pixel's crop?

4.Classification Outcome: Based on the rules, fires were classified into four types:

Maize_Straw_Burning, Wheat_Straw_Burning, Cropland_Fire_Other and Non_Cropland_Fire.

Task 2: What’s the Percentage?

--- 目标: 筛选黑龙江省火点并匹配到县... ---

假定文件在当前工作目录: C:\Users\46400\Desktop\GKG\上课\seminar 1\CNGF5020 (L01) - Environmental Modeling and Big Data Analytics\作业\Mini Project I\Mini Group Project I Data

正在加载 modis_2010_2019_cleaned.csv...

成功加载 751014 条高质量火点。

正在加载 CHN_County.shp...

成功加载 2860 个县级多边形。

成功从地图中筛选出 128 个黑龙江省的县/区。

--- 正在执行空间连接...(这可能需要1-2分钟) ---

--- 空间连接完成! 耗时 0.72 秒 ---

成功! 在 751014 个总火点中,

有 135512 个火点位于黑龙江省境内。

Task 2: What’s the Percentage?

```
============================================================
--- 火灾分类结果统计 ---
fire_type
Non_Cropland_Fire       105082
Cropland_Fire_Other      27540
Maize_Straw_Burning       2807
Wheat_Straw_Burning         83
Name: count, dtype: int64
============================================================
```

Task 2: What's the Percentage?

Calculation -- Based on the results from Script:

Number of Maize_Straw_Burning fires = 2,807

Number of Wheat_Straw_Burning fires = 83

Total Straw Burning Fires = Maize + Wheat = 2,807 + 83 = 2,890

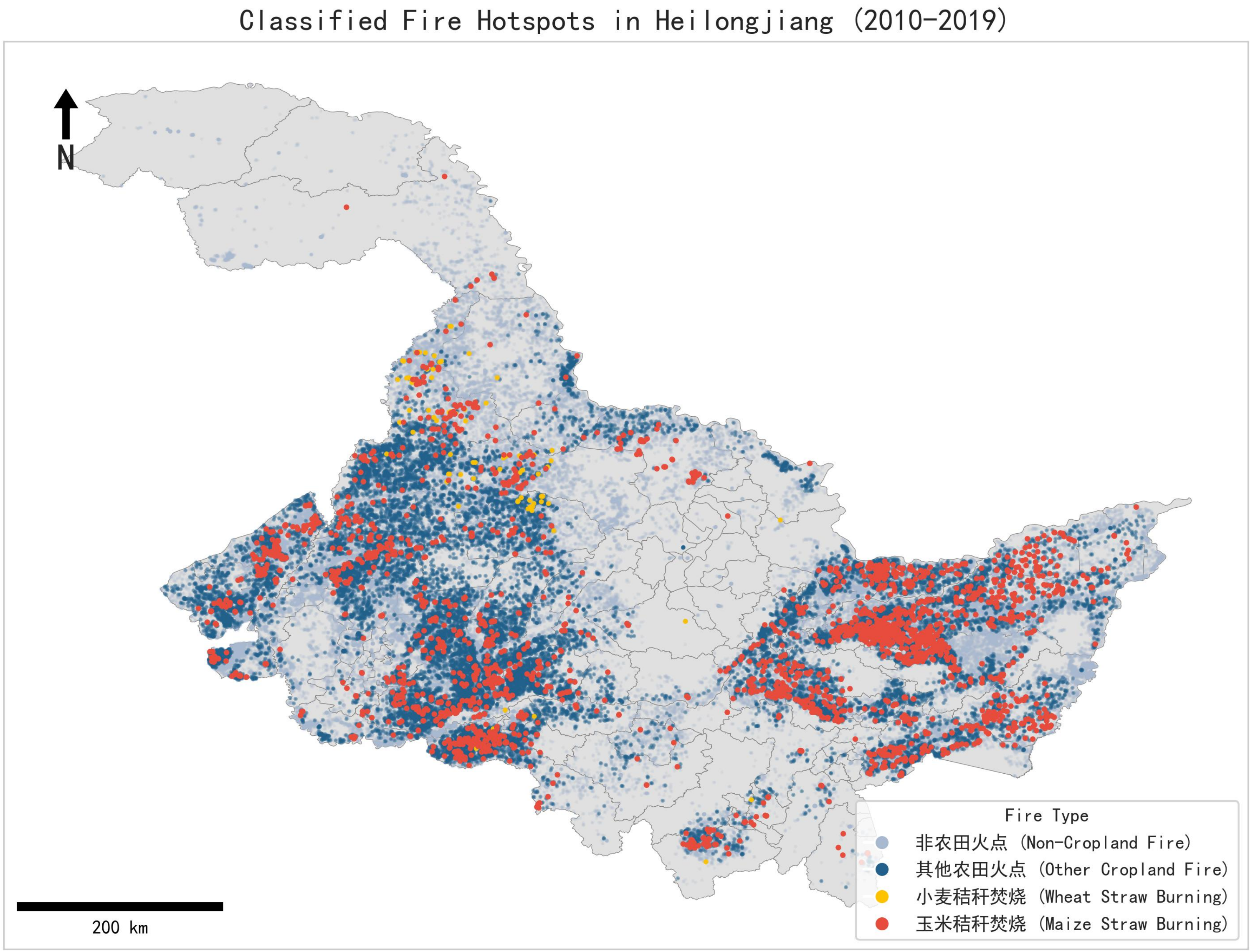
Total number of high-confidence fires in Heilongjiang = 135,512

Percentage = (Total Straw Burning Fires / Total Fires in Heilongjiang) * 100%

= (2,890 / 135,512) * 100% ≈**2.13%**

*We classified approximately **2.13%** of fires as likely straw burning (post-harvest burning of corn and wheat).

Task 2: What's the Percentage?



Task 3: Long Term Spatiotemporal Trends

--- 目标: 分析黑龙江省各县秸秆焚烧的长期趋势... ---

警告: 无法检测 __file__。假定文件在当前工作目录: C:\Users\46400\Desktop\GKG\上课\seminar 1\CNGF5020 (L01) - Environmental Modeling and Big Data Analytics\作业\Mini Project I\Mini Group Project I Data

正在加载 modis_heilongjiang_classified_fires.csv...

成功加载 135512 条已分类的火点。

--- 正在筛选和统计秸秆焚烧火点... ---

总共筛选出 2890 个秸秆焚烧火点。

成功创建了县级年度火点统计透视表。

已将统计数据表保存到: C:\Users\46400\Desktop\GKG\上课\seminar 1\CNGF5020 (L01) - Environmental Modeling and Big Data Analytics\作业\Mini Project I\Mini Group Project I Data\task3_county_straw_burning_counts.csv

--- 正在绘制县级趋势热力图... ---

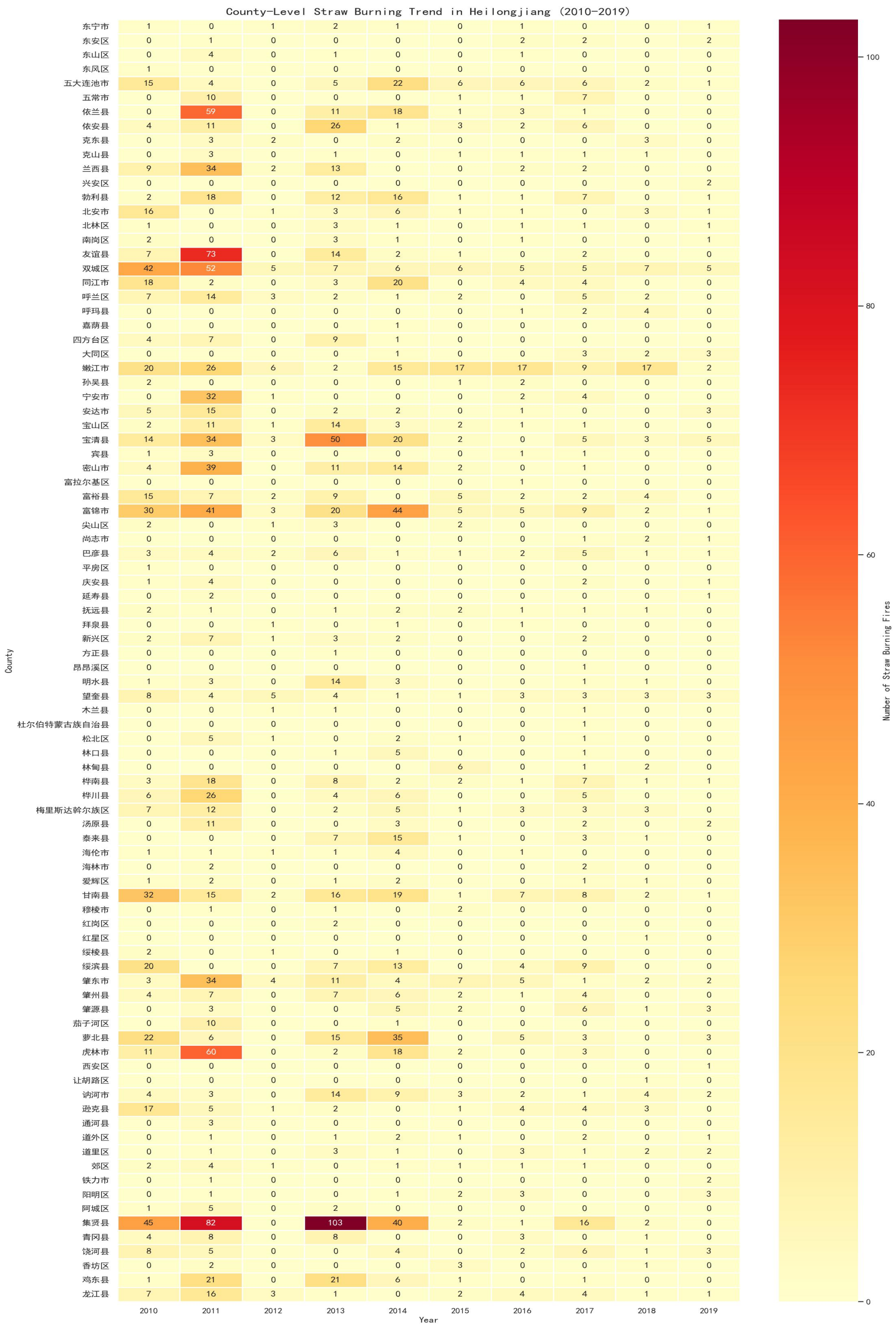
成功! 已将县级趋势热力图保存到: C:\Users\46400\Desktop\GKG\上课\seminar 1\CNGF5020 (L01) - Environmental Modeling and Big Data Analytics\作业\Mini Project I\Mini Group Project I Data\task3_county_straw_burning_trend_heatmap.png

Task 3

Signs of Reduction:

Straw burning activity shows signs of reduction from its peak levels in the later part of the decade, but not necessarily a continuous decline over the entire 10–year period relative to 2010.

The trend is more like **an increase followed by a decrease**.

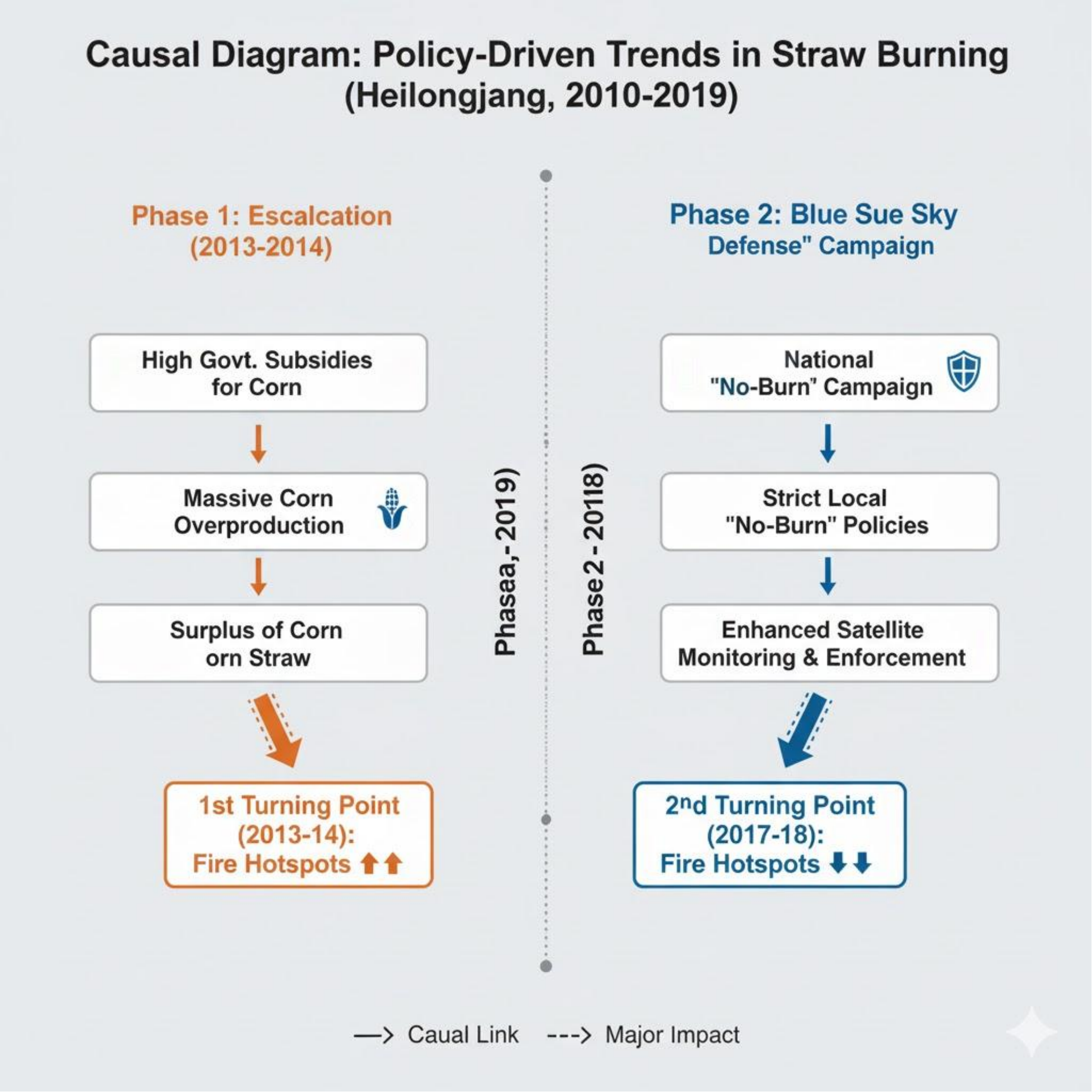


Turning Points and Shifts:

First Turning Point (Increase): Around **2013–2014** appears to be a turning point where the intensity and frequency of straw burning escalated notably before this period.

Second Turning Point (Decrease): Around **2017–2018** seems to mark another turning point where activity began to decrease substantially across numerous counties compared to the peak.

Task 3: Long Term Spatiotemporal Trends



Not only what and why

but also **HOW**

to make the processing of straw

GOOD FOR FARMERS, ENVIRONMENT AND ALL OF US !

MANY THANKS!

Let’s leave the rest of the time to Jimmy.

Task 4

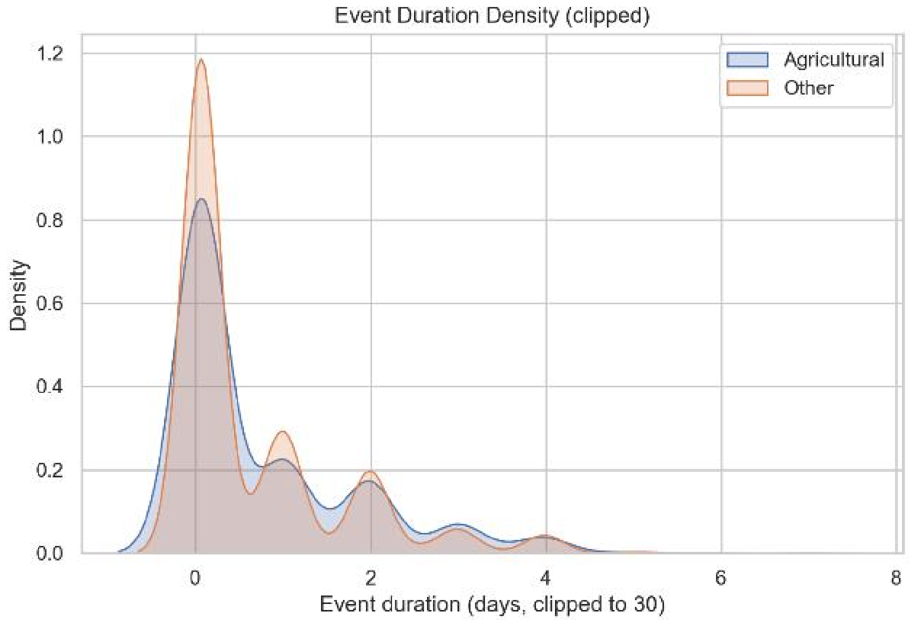
Agricultural fire definition: All fires occurring on farmland, including Maize_Straw_Burning + Wheat_Straw_Burning + Cropland_Fire_Other

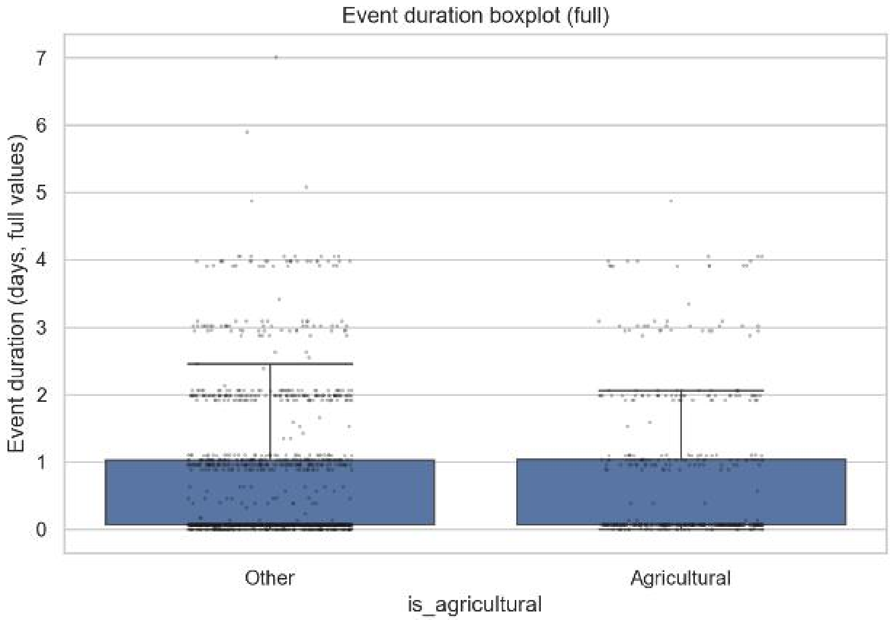
Metrics: Event duration, DOY, Hour, FRP.

Highlights: Using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method in unsupervised learning within machine learning, which combines spatial and temporal analysis, DBSCAN clustering is used to estimate the "event duration" (not a single point, but the fire event level).

A real fire often spreads spatially and lasts for hours or days. Therefore, simply looking at the distribution of points ignores the overall nature of the event.

Task 4





Both are concentrated in the 0—2 day range, both fire events are very short-lived.

The agricultural fire curve is slightly skewed to the right, they may last slightly longer on average.

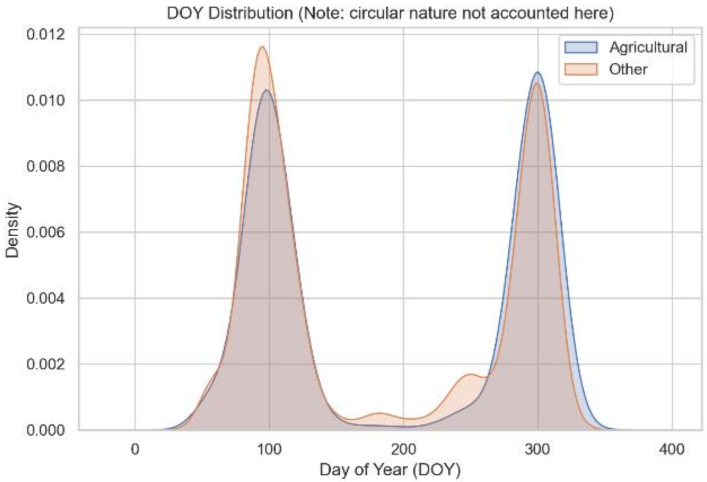
However, the overall difference between the two is not significant, with similar peak positions and most events ending within one day.

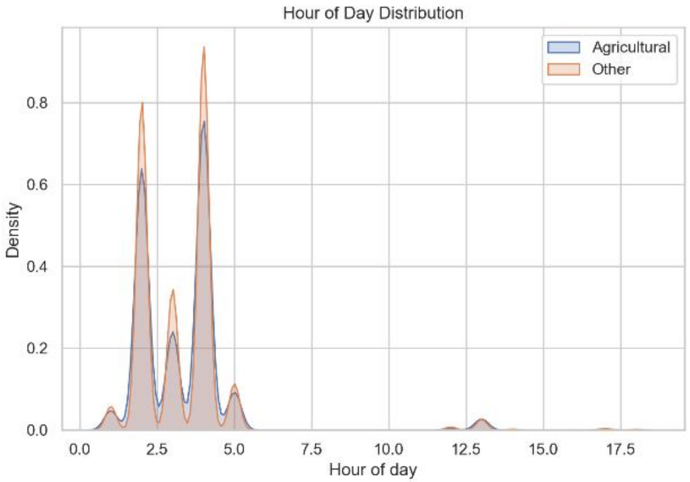
The median duration for both agricultural fires and other fires is close to 1 day.

Agricultural fires have slightly longer duration time, a smaller number of agricultural fire events lasted longer.

The distributions of both types of fires highly overlap, suggesting that the overall differences are not significant.

Task 4





The two significant peaks of agricultural fire occurred at 100 days (April) and 290 days (October), precisely corresponding to the sowing and harvesting periods of corn and wheat in Heilongjiang Province.

This high concentration and peak indicate a clear seasonal pattern in agricultural fires.

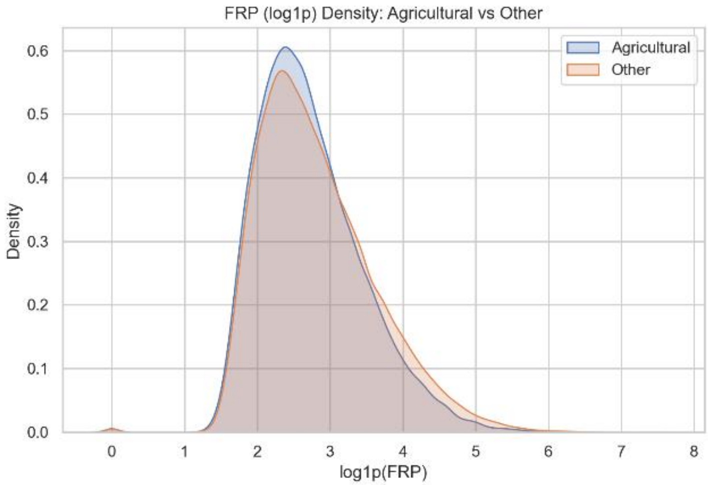
However, other fires also exhibited a similar distribution, albeit a more gradual one.

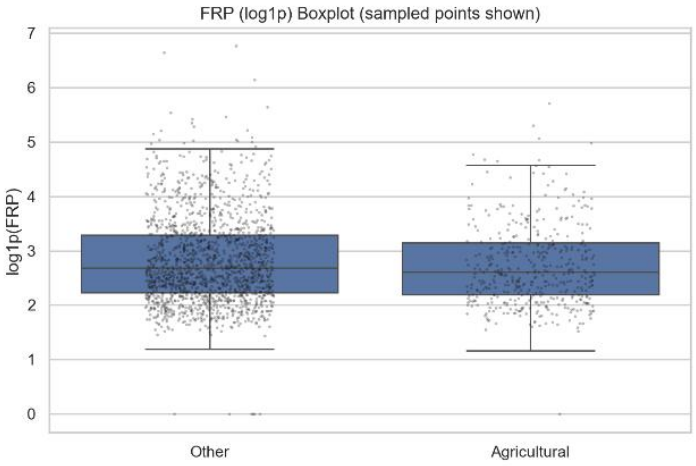
Both types of fires showed significant peaks between 2 and 5 AM.

This primarily reflects the timing characteristics of MODIS satellite observations (Terra/Aqua satellites pass overhead during this period) and therefore cannot fully represent human ignition behavior.

The temporal distributions of agricultural fires and other fires are very similar, because both are largely controlled by satellite observation time.

Task 4





The two curves are very similar in shape, both exhibiting a right-skewed distribution , indicating that most fire spots have low FRP , with only a few fire spots being extremely intense.

Agricultural fire spots have a slightly higher density peak, concentrated at log(FRP) ≈ 2.3 (i.e., FRP ≈ 10 MW), indicating that the fire intensity of agricultural fire spots is relatively more concentrated.

Other fire spots have a slightly wider right tail, suggesting that some non-agricultural fire spots (such as industrial fires and forest fires) may have higher FRP, burning more intensely or lasting longer.

Overall, the two types of fire spots do not differ significantly in FRP distribution, but "other fires" are more prone to extreme high-intensity events.

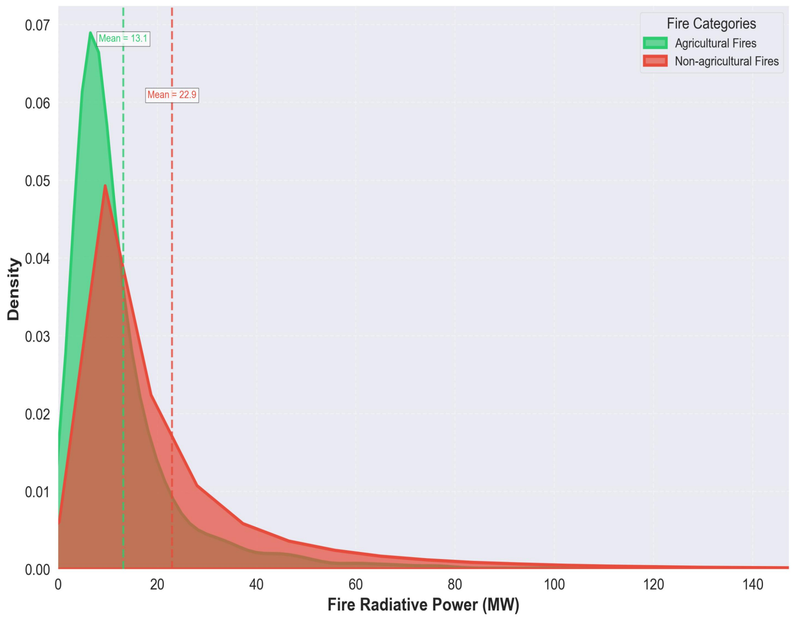
The median and interquartile ranges of the two boxes almost overlap, indicating a similar overall FRP distribution between the two fire types.

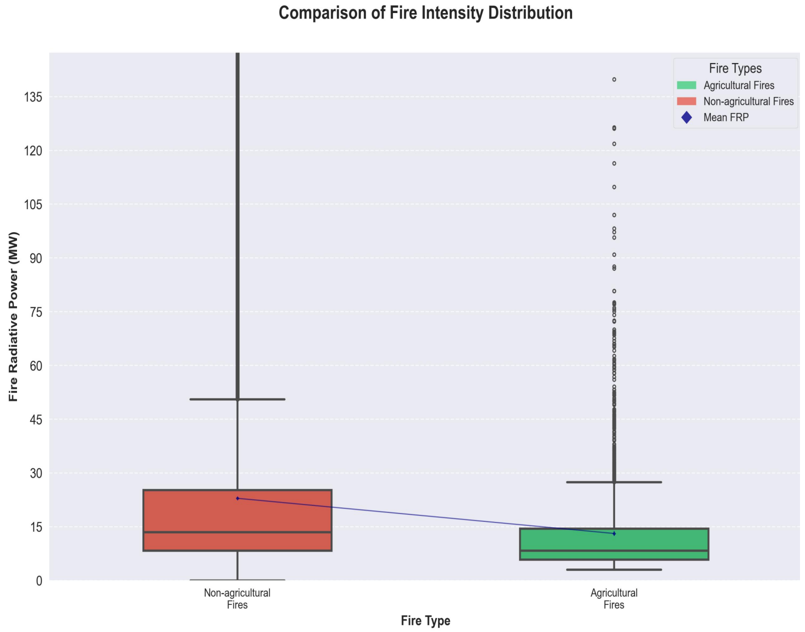
The "Other Fires" box is slightly higher, suggesting a higher upper limit for fire intensity.

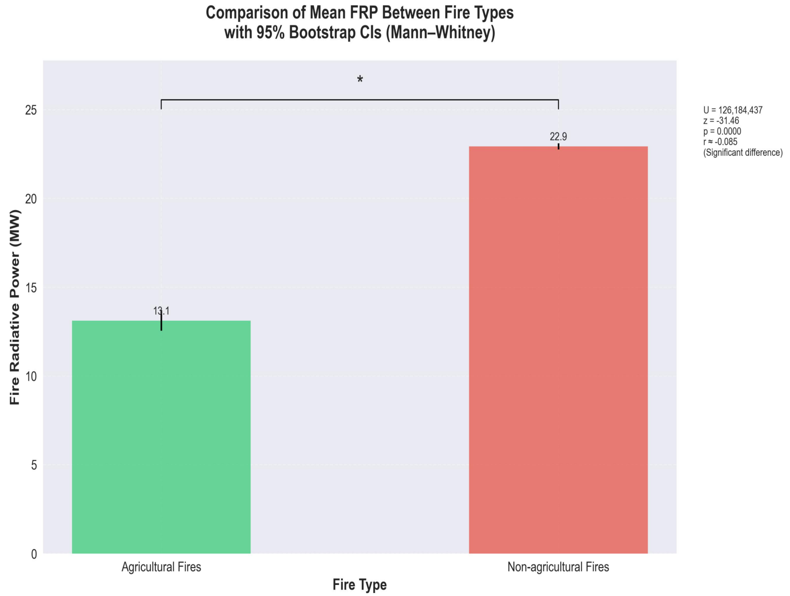
No significant systematic shift is observed, implying that the average difference in fire intensity between agricultural fires and other fires is not significant.

Challenge 1

Agricultural fire definition: Maize_Straw_Burning + Wheat_Straw_Burning.







The both fire intensities are concentrated at relatively low levels. Compared to non-agricultural fires, agricultural fires are more clearly concentrated at low intensities, while non-agricultural fires have more extreme values.

We performed a U-test on the statistical results, and the results showed that $p < 0.05$. This indicates that the intensity of "non-agricultural fires" is statistically significantly higher than that of "agricultural fires".

Challenge 1

Descriptive statistics (selected)

	Agricultural	Non-agricultural
count	2890.0	132622.0
mean	13.12	22.93
std	16.16	31.9
min	3.0	0.0
25%	5.8	8.3
50%	8.3	13.5
75%	14.48	25.2
max	315.4	1824.4

The result shows that the fire intensity between the two group has significant difference.
So agricultural may be not the source of major intense buring events, and wildfires can cause much larger damage compared to agriculture fires.

Challenge 2

In fire classification, we used three filtering methods: global (confidence) filtering, spatial filtering, and time window filtering to distinguish farmland fires and farmland fires caused by straw burning from other fires.

The sources of uncertainty and the potential errors they cause are as follows:

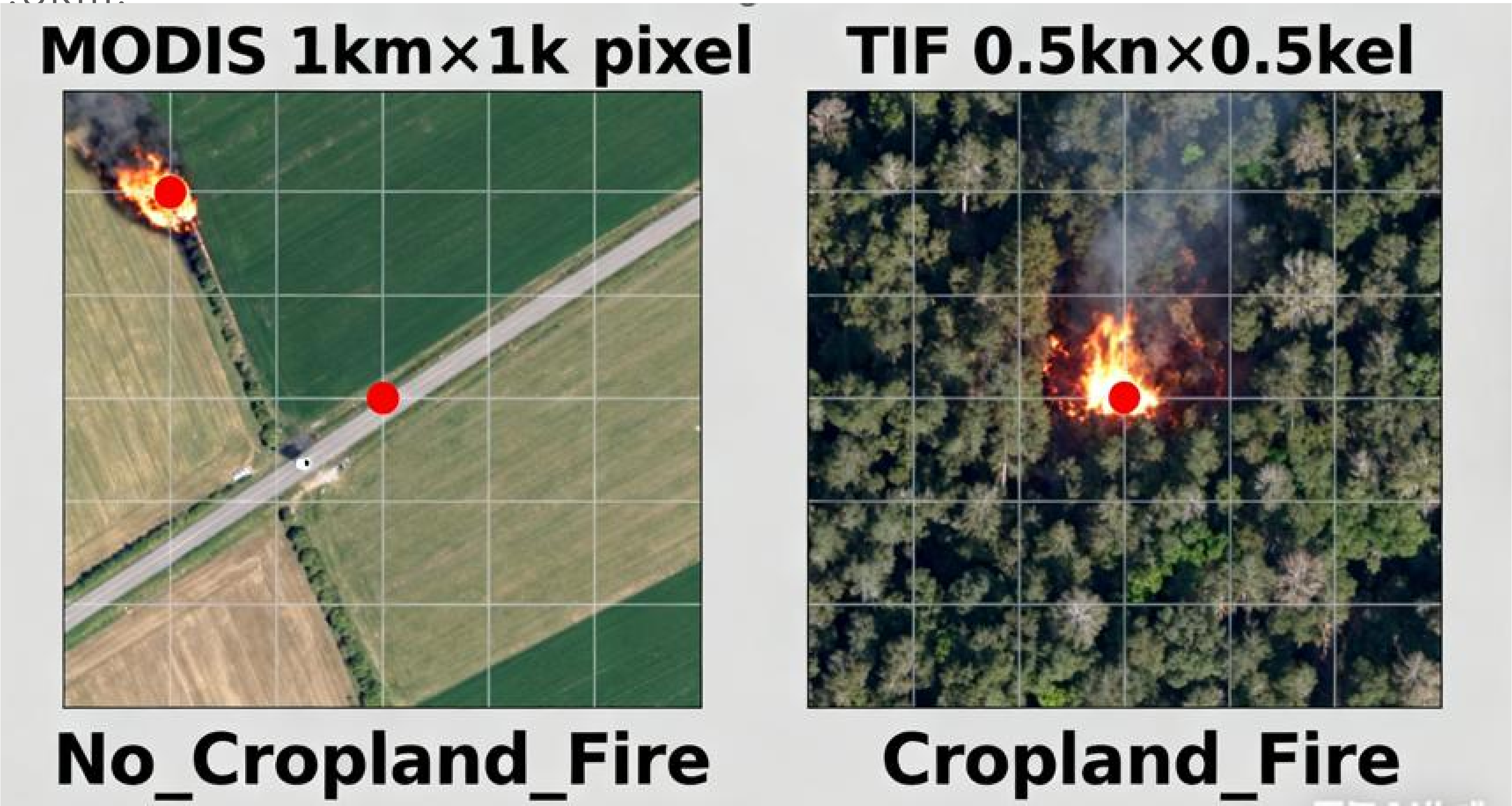
1. TIF farmland maps are inaccurate, misidentifying grassland as cornfields and inflating the number of farmland fires. Maps are also time-sensitive; farmers rotate crops, so this year's corn might be shown as last year's soybeans, reducing the number of farmland fires.

2. Burning times are determined by harvest cycles, meaning farmers' burning activities may occur outside the [mature DOY – 15, mature DOY + 30] time window. This could be due to delays caused by rain or policy changes.

3. MODIS data is affected by cloud cover and smoke. During large-scale burning, dense smoke may obscure fires in other locations, leading to an underestimation of the true number of fires.

Challenge 2

4. The resolution of MODIS and farmland maps is mismatched. The resolution of MODIS fire points is 1km*1km, while TIF farmland may have a higher resolution of 0.5km*0.5km.



Challenge 2

Therefore, we propose using soil moisture data to improve classification accuracy. The core logic is : Straw burning is a deliberate, planned fire aimed at removing dried crop residue.

Therefore, it almost always occurs when the surface and topsoil are dry. High soil moisture (e.g., immediately after rain or during irrigation) inhibits straw burning. Low soil moisture (e.g., after harvest, in sunny, dry weather) is a necessary condition for straw burning.

We plan to use daily soil moisture data that matches the resolution of the farmland map, and process it as follows:

1.Calculate the monthly average soil moisture for each TIF pixel between 2010 and 2019. Then, for each fire, calculate its SM anomaly = (daily SM value) − (monthly average SM value). A negative anomaly indicates drier conditions than usual (suitable for incineration); a positive anomaly indicates wetter conditions than usual (unsuitable for incineration). An extreme negative anomaly value is set, such as SM anomaly < -0.08.

Challenge 2

2. Process Non_Cropland_Fire, Maize/Wheat_Straw_Burning, Cropland_Fire_Other:

- Non_Cropland_Fire (SM anomaly):
  - Negative → Probable_Straw_Burning (Spatial_Error)
  - Extremely negative → Forest wildfire
  - Positive → Industrial heat/data noise
- Maize/Wheat_Straw_Burning (SM anomaly):
  - Negative → Confirmed_Straw_Burning
  - Positive → Anomalous_Cropland_Fire
- Cropland_Fire_Other (SM anomaly):
  - Negative → Probable_Straw_Burning (Time_Window_Error)
  - Positive → Retain Cropland_Fire_Other

3.new estimate will be the sum of (Confirmed_Straw_Burning + Probable_Straw_Burning (Time_Window_Error) + Probable_Straw_Burning (Spatial_Error)).

MANY THANKS AGAIN FOR LISTENING!

GOOD NIGHT!