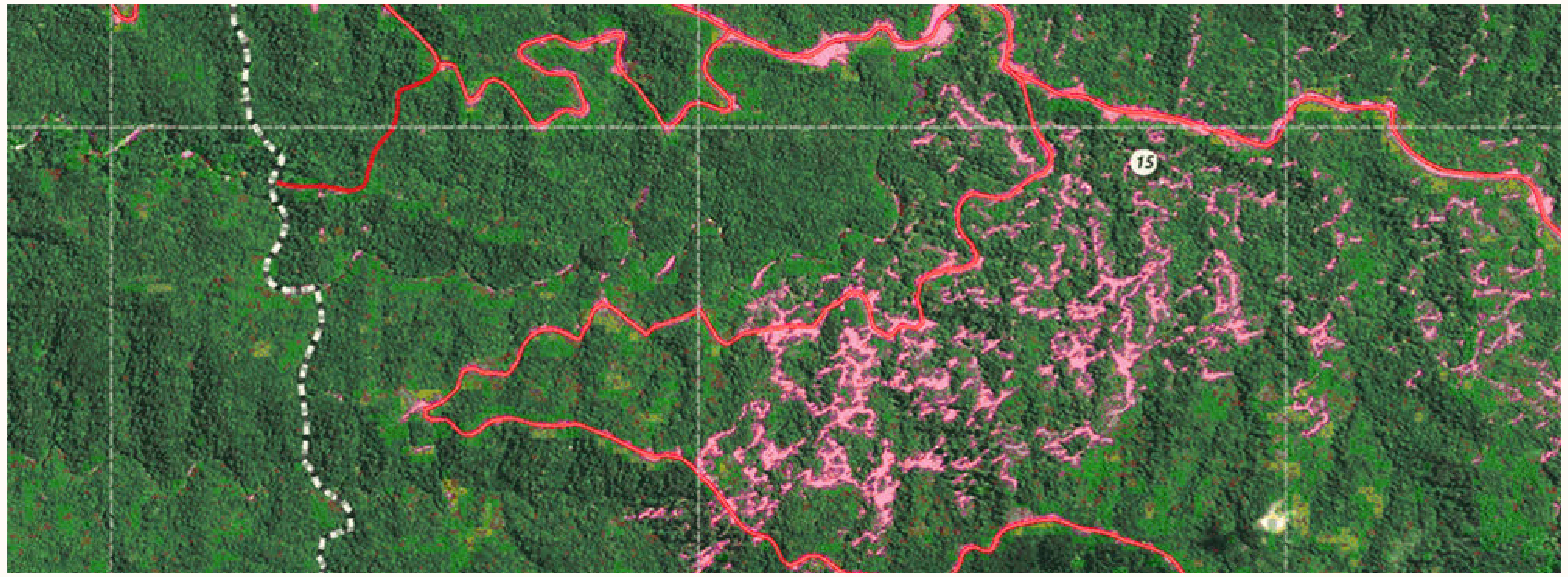
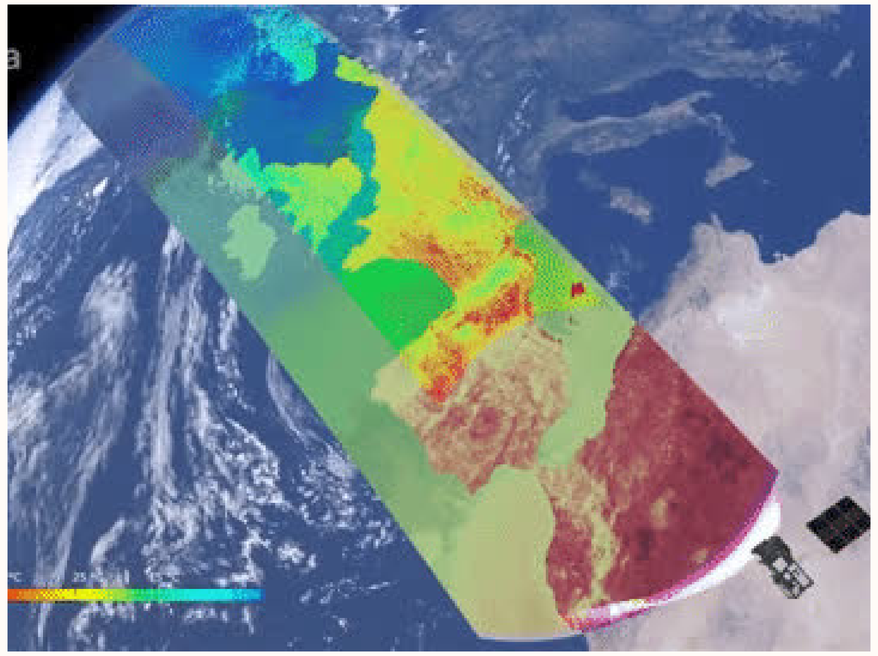
SICKLE: A MULTI-SENSOR SATELLITE IMAGERY DATASET ANNOTATED WITH MULTIPLE KEY CROPPING PARAMETERS

EXTENDING SICKLE: ZERO-SHOT INFERENCE ON PADDY FIELDS IN ANDHRA PRADESH

REMOTE SENSING

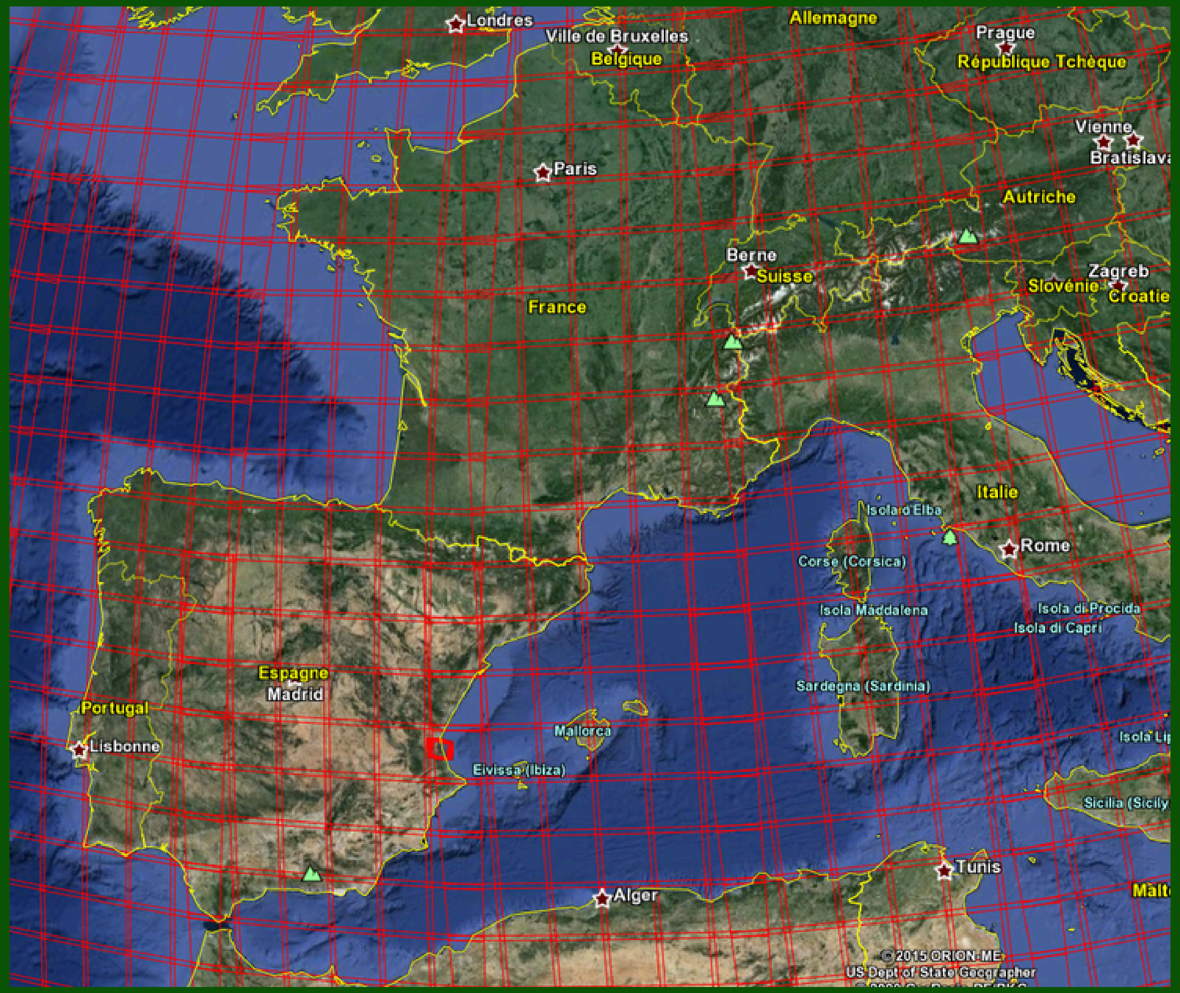
- Can generate "policy relevant" data for "large and inaccessible" areas in a cost effective manner
- Increased applicability with integration of modern machine learning paradigms





REMOTE SENSING FOR AGRICULTURE

- Wide applications for agriculture!
- However, collecting field data for agricultural applications involves conducting ground based surveys which is expensive and laborious
- This is why there is a lack of labelled datasets despite abundance of satellite images!

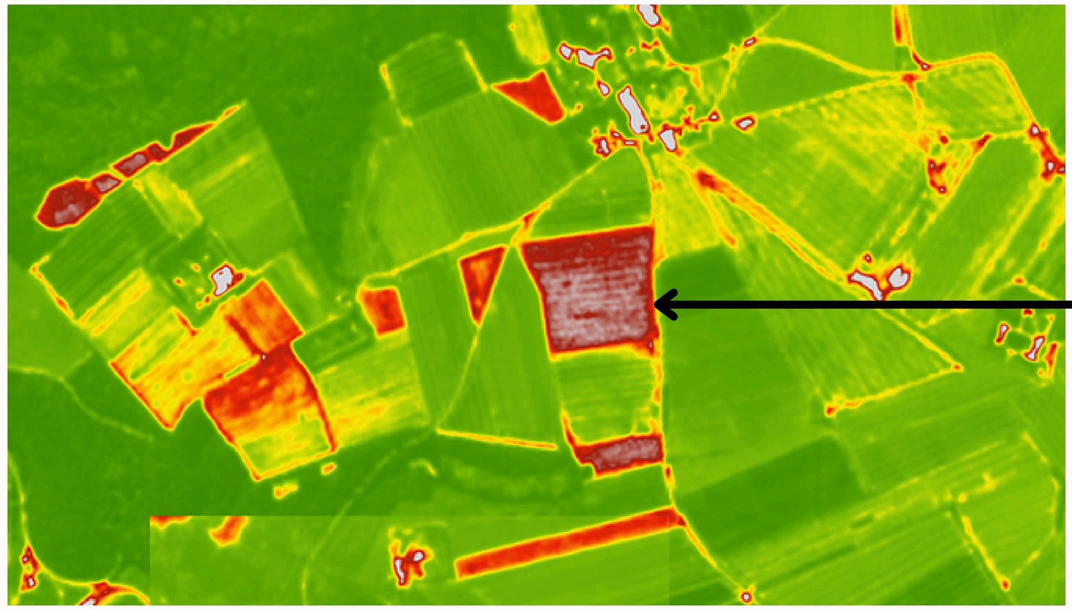


ISSUES FOR CREATING LABELLED DATASET

- High cost of conducting surveys
- Difficult to survey a large geographical region by physical interacting with farmers
- Reluctance of farmers to share data
- Complex satellite manual annotation process
- Satellite images are organized as big tile images covering multiple sq km of geo areas
- Due to their high resolution make it hard to acquire their respective time series data. This also make them harder to process

WHERE PREVIOUS WORKS LACKED

- Most works focused on a single task, while monitoring cropping pattern is a combination of multiple connected tasks. No work provides multiple cropping params for the same set of crops!

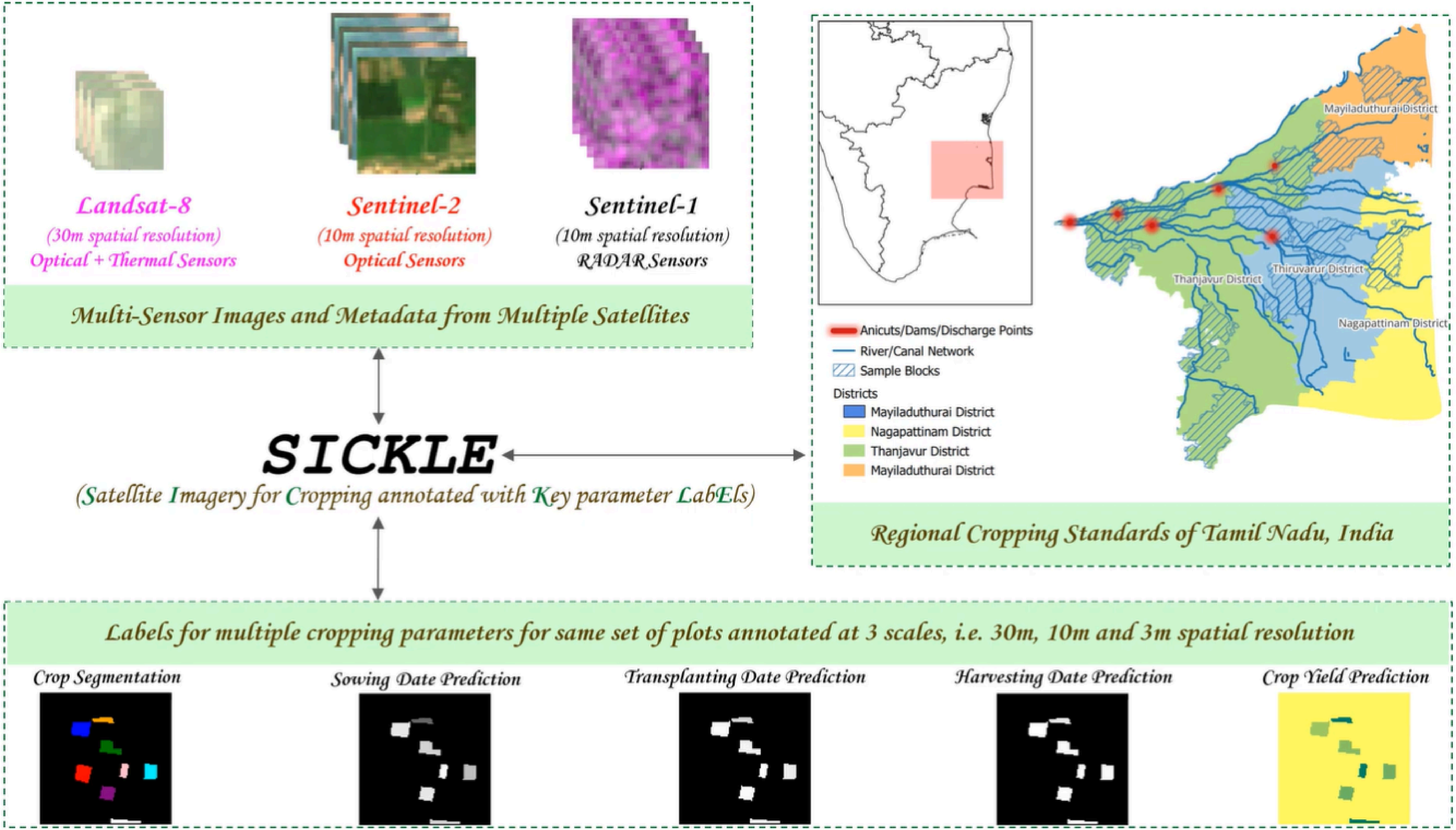


Crop Type: Rice
Productivity: ?????

- Farmers can grow different crops with different growing seasons. No works includes information about the growing season of different crops!

Crop Type: Rice
Growing Season: March-May

Crop Type: Maize
Growing Season: May-June


Landsat-8
(30m spatial resolution)
Optical + Thermal Sensors
Sentinel-2
(10m spatial resolution)
Optical Sensors
Sentinel-1
(10m spatial resolution)
RADAR Sensors
Multi-Sensor Images and Metadata from Multiple Satellites
SICKLE
(Satellite Imagery for Cropping annotated with Key parameter LabEls)
Mayiladuthurai District
Thanjavur District
Thiruvarur District
Nagapattinam District
Anicuts/Dams/Discharge Points
River/Canal Network
Sample Blocks
Districts
Mayiladuthurai District
Nagapattinam District
Thanjavur District
Mayiladuthurai District
Regional Cropping Standards of Tamil Nadu, India
Labels for multiple cropping parameters for same set of plots annotated at 3 scales, i.e. 30m, 10m and 3m spatial resolution
Crop Segmentation
Sowing Date Prediction
Transplanting Date Prediction
Harvesting Date Prediction
Crop Yield Prediction

SICKLE

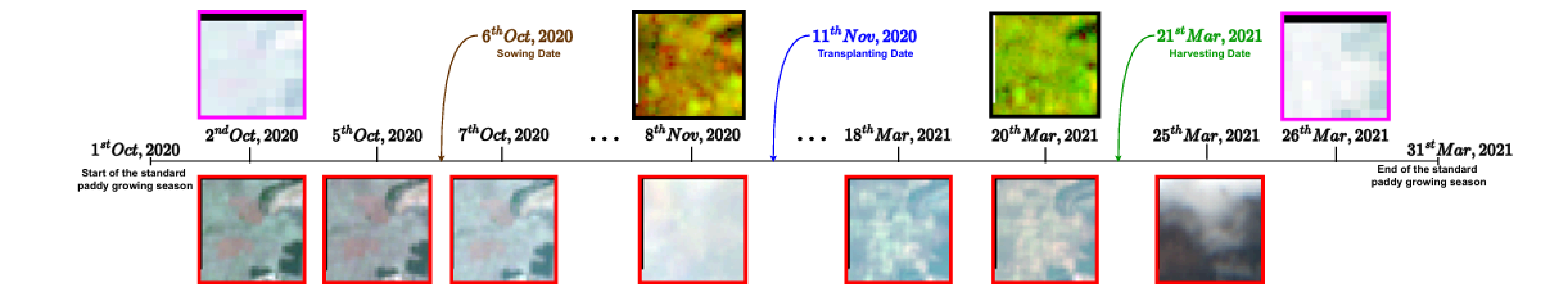
SATELLITE IMAGERY FOR CROPING ANNOTATED WITH KEY PARAMETER LABELS

- Multi-sensor (S1, S2, L8), time-series satellite imagery
- Region: Cauvery Delta, Tamil Nadu
- Annotated with 6 key tasks:
  - Crop type, segmentation
  - Phenology (sow, transplant, harvest)
  - Crop yield
- 2370 samples, 388 plots
- Tasks: Crop type prediction (Classification), Crop Phenology (Regression), Crop Yield (Regression)

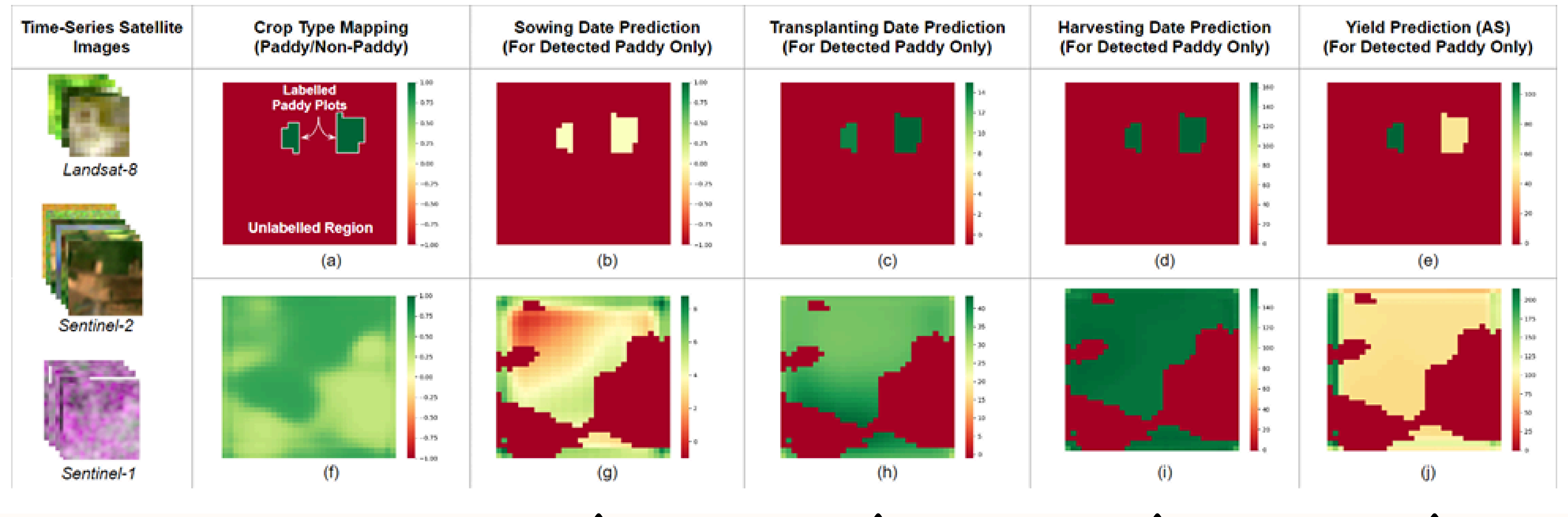
UNDERSTANDING TIME SERIES DATA

- The length of "Regional standard growing season of paddy" is considered as the duration for creating the time series data

- Regional standard growing season is available even in the absence of information of crop's actual growing season (often the case in real world setting)



OUTPUTS OBTAINED FROM U-NET 3D (FUSION)



Ground truth annotations

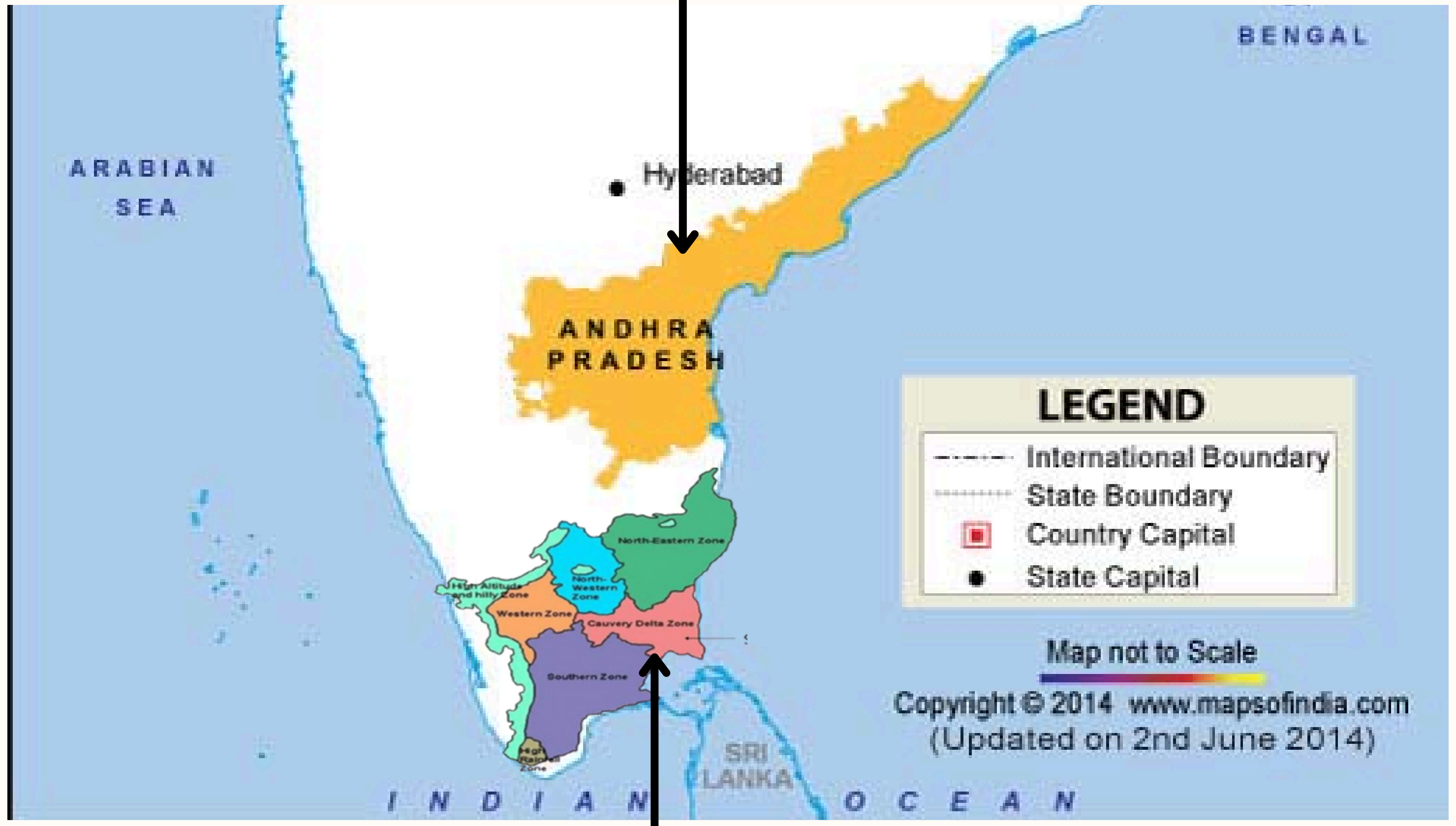
Generated output maps

All non paddy predicted pixels are masked out here

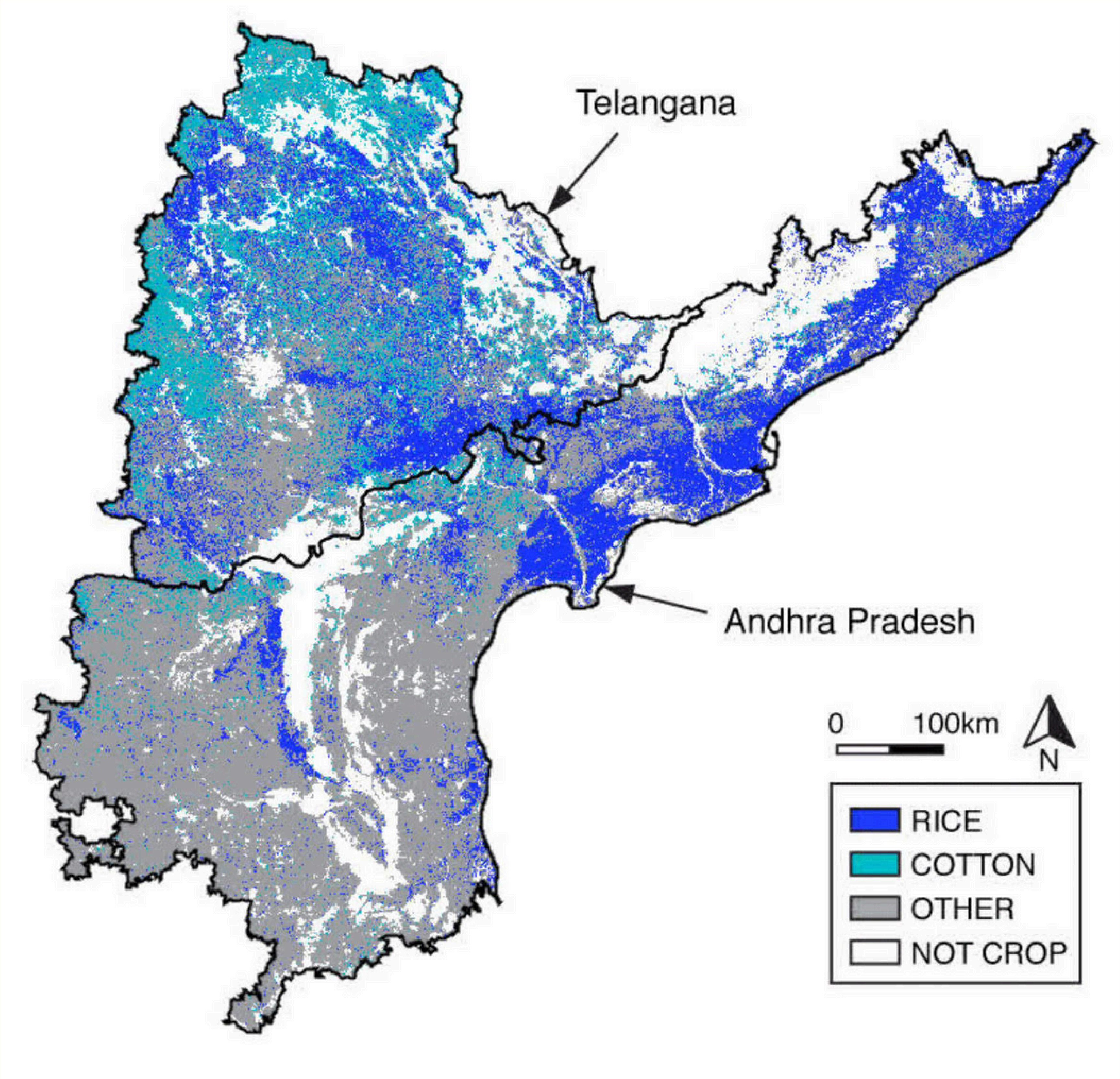
MOTIVATION

- Agriculture in India is regionally diverse — models must generalize beyond one state.
- Existing datasets are scarce, especially those annotated for key cropping parameters.
- SICKLE introduced a strong baseline, but it was limited to Tamil Nadu.
- Can models trained there generalize to Andhra Pradesh?
- Andhra Pradesh is a very contrasting region from Cauvery Delta, With arid regions while original area of study is a humid place.

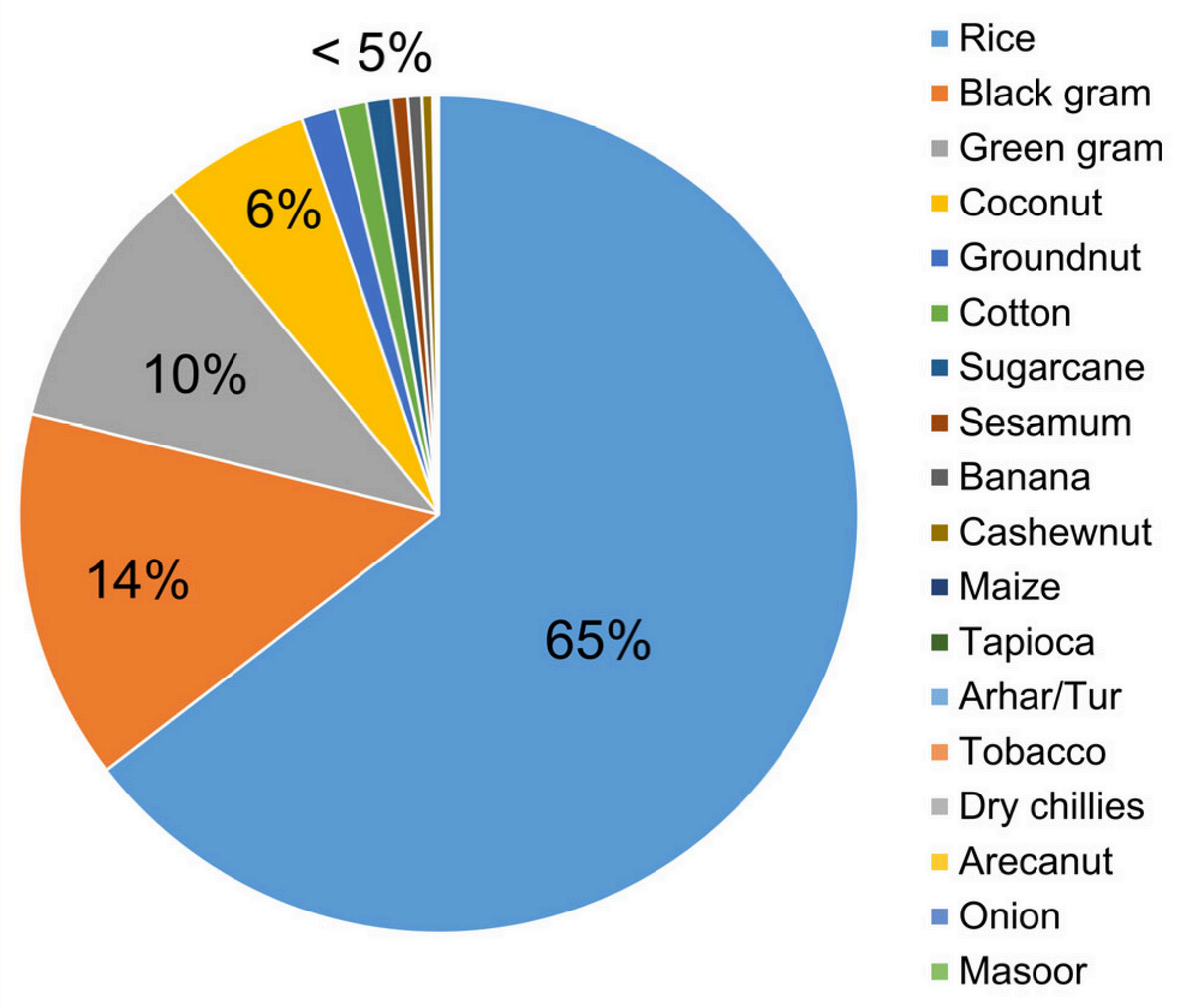
Our Study area:
Andhra pradesh State



SICKLE Study area:
Cauvery Delta Region (RED)



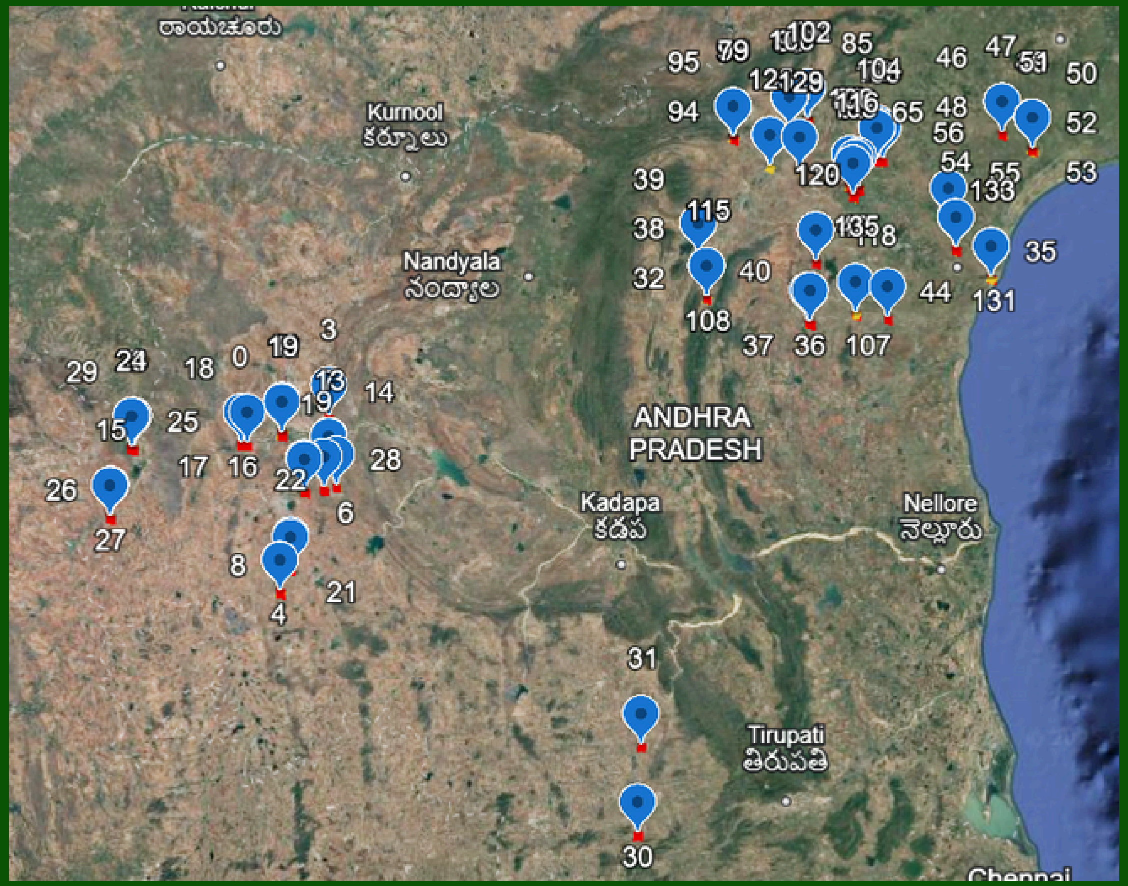
Andhra Pradesh Crop Types (Our study area)



Cauvery Delta Crop Types

OUR CONTRIBUTION

- Created a new dataset of paddy plots in Andhra Pradesh
- Used CSISA_IND_LDS_Rice_2018_Data[1] to obtain plot level info in Andhra Pradesh
- Manually Annotated 150+ plots using Google Earth Pro
- Acquired corresponding Sentinel-1 imagery
- Applied SICKLE's pretrained Sentinel-1 model for inference only
- **Goal: Zero-shot generalization to new geography**



ORIGINAL INFERENCES FROM THE PAPER

Task	Metric	L8	S2	S1	Fusion
Crop Type (SI)	IoU (%)	47.73% ± 1.77%	54.87% ± 3.08%	**64.35% ± 4.82%**	-
Crop Type	IoU (%)	56.04% ± 5.84%	78.12% ± 3.48%	**81.77% ± 6.60%**	**81.07% ± 5.77%**
Sow Date	MAE (days)	2.66 ± 0.961	**2.30 ± 0.611**	3.61 ± 0.898	**2.33 ± 0.639**
Transplant Date	MAE (days)	**6.20 ± 1.030**	6.36 ± 2.164	7.23 ± 0.779	**6.16 ± 1.770**
Harvest Date	MAE (days)	**9.86 ± 0.736**	**8.83 ± 1.520**	10.08 ± 0.561	10.75 ± 3.389
Crop Yield (SI)	MAPE (%)	**46.74% ± 3.82%**%	60.44% ± 14.50%	**48.35% ± 7.64%**	-
Crop Yield (RS)	MAPE (%)	**54.00% ± 9.67%**	72.38% ± 8.74%	71.81% ± 17.27%	**70.35% ± 13.75%**
Crop Yield (AS)	MAPE (%)	**59.38% ± 14.75%**	73.59% ± 9.81%	65.66% ± 16.24%	**64.56% ± 13.77%**

Benchmarked Tasks include, Binary Crop Type Mapping, Sowing Date Prediction, Transplanting Date Prediction, Harvesting Date Prediction and Crop Yield Prediction

OUR TASKS AND APPORACH

Accuiring Qualitative & Quantitative Data

CSISA's survery data was the main dataset for conducting inferences.

This dataset was part of the "Large-scale data of crop production practices applied by farmers on their largest rice plot during 2018 in eight Indian states" paper.

Accuiring Image Data

Image data or satellite data was aqurired using Sentinel Hub, namely with 3 different sattelites, Sentinel-1, Sentinel-2, Landsat-8.

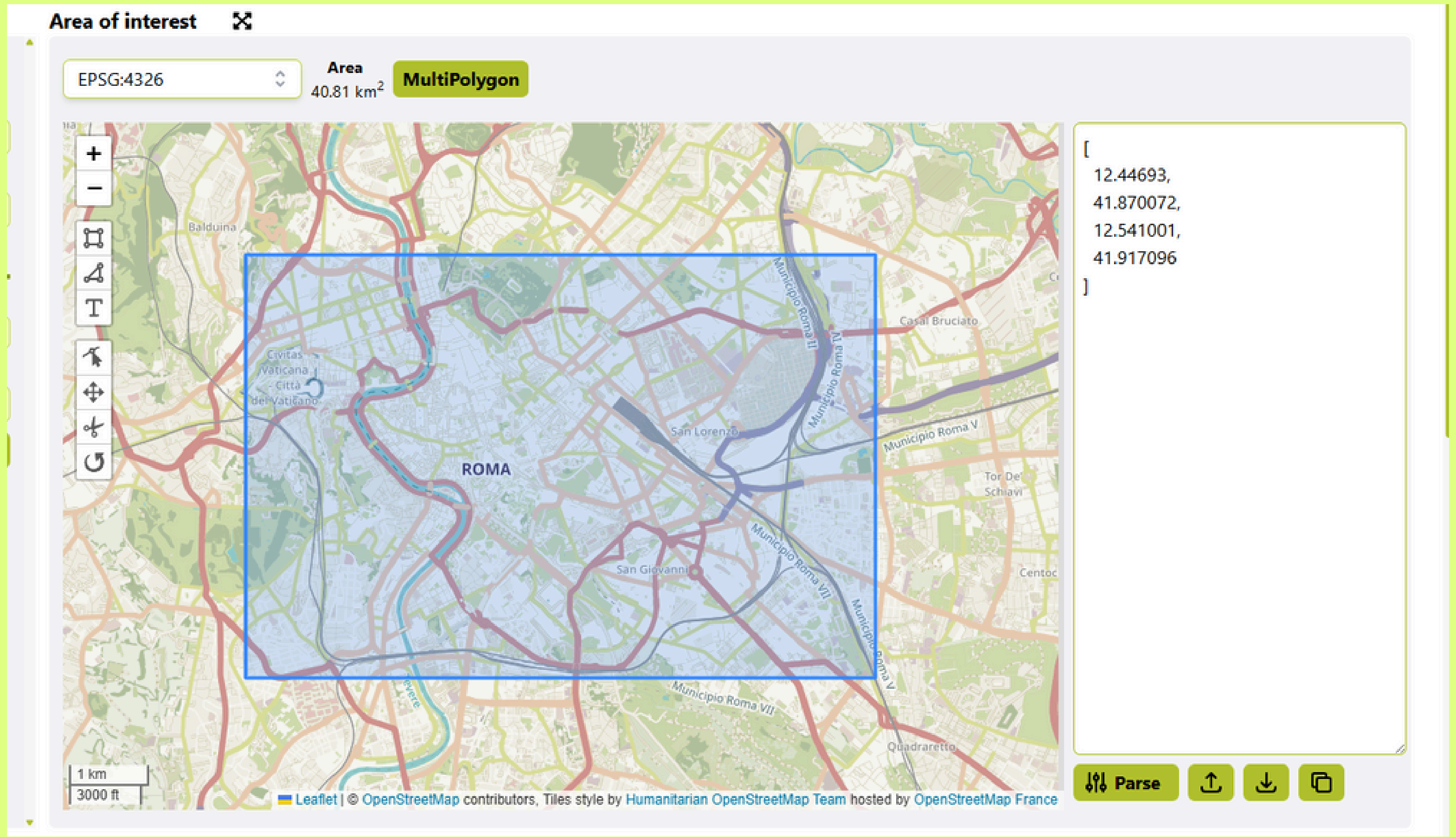
Products Listed down:
Sentinel-1/GRD
Sentinel-2/L2A
Landsat-8/L2

METHODOLOGY (IMAGERY)

We used the SentinelHub Api to retrieve our plot images. After plotting out the areas, we retrieved various GeoJSON files of different plots, such as bounding boxes, and a satellite image that corresponds to a particular area.

The main ideology was to generate plots for each farmland first and then retrieve the shapefiles. Generation of bounding boxes and shape files is mentioned in **Later Slides**.

- First, we received the plots in KML files. Which is explained later how.
- Converted into GEOjsons and then used as a bounding box or a filter of the entire state to get the correct data.

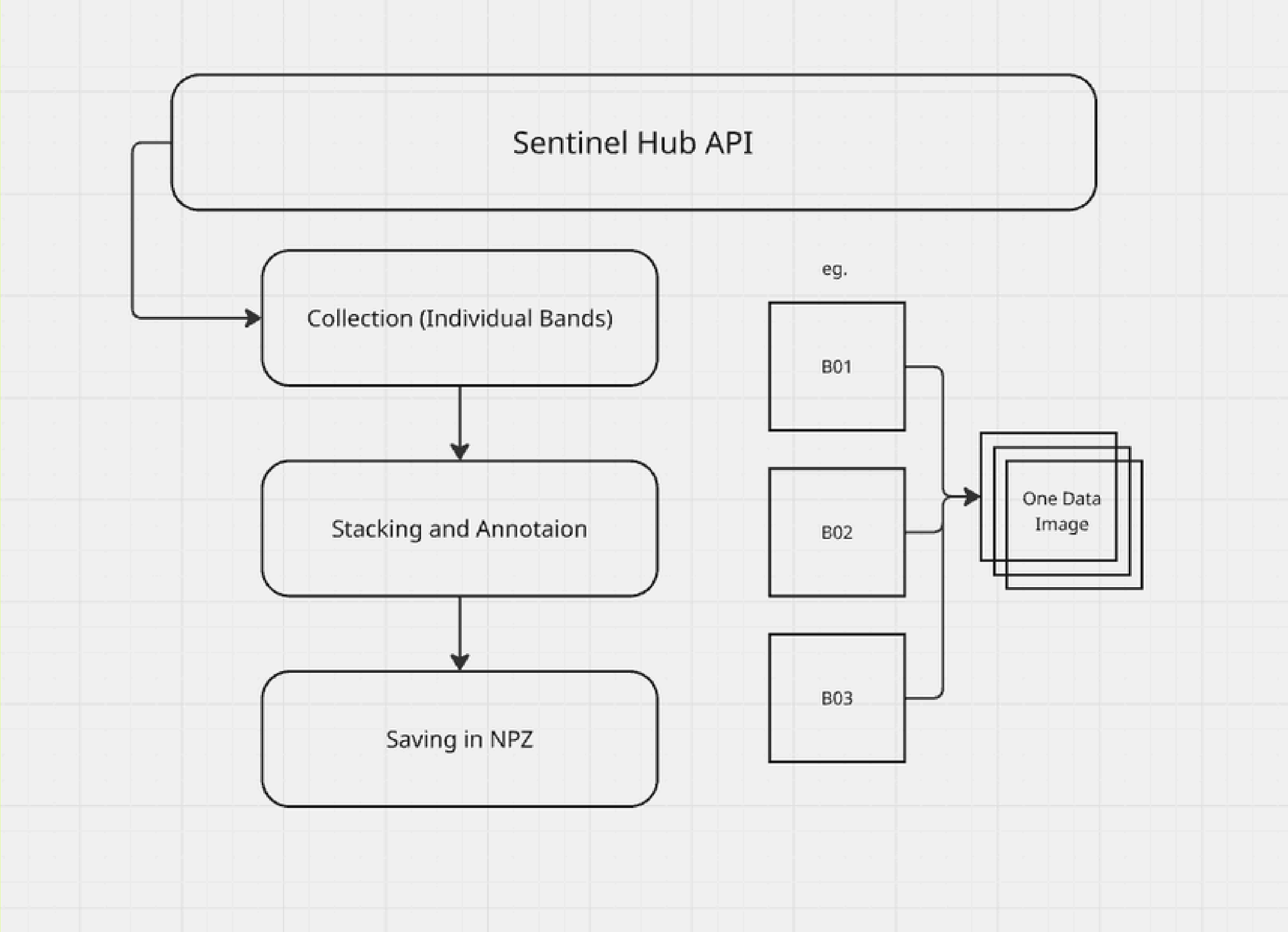


SentinelHUB requests builder shows how a bounding box is supposed to work, we are retrieving the area defined within.

METHODOLOGY (CONTINUED)

The Retrieval and the pre-processing Pipeline works as follows:

- Data Retrieval
  - Fetch each band separately for every bounding box from the source.
- Band Stacking & Annotation
  - Stack the bands into a single multi-layer raster per acquisition.
  - Assign each layer a unique key (e.g. "SR_B1", "SR_B2", ...).
- Directory Organization
  - Create one folder per bounding box.
  - Within each, maintain subdirectories for each satellite and date range (e.g. 2018-08-01_to_2018-08-31).
- Quality Control & Masking
  - Apply cloud/shadow masks (e.g. QA bands or external algorithms) to filter out bad pixels.
  - Log quality metrics (e.g. cloud cover %) for each time step.
- Array Export & Metadata Logging
  - Convert each stacked raster to a NumPy array and save with its annotation keys.
  - Record metadata (timestamp, spatial resolution, CRS) alongside each array for downstream analysis.



SATELLITE DATA & KEY BANDS

Sentinel-1/GRD: Radar type with SAR Sensors. It measures Surface texture, moisture, and structures, Works in all weather and day/night. In this use case only two bands are used VV and VH.

Sentinel-2/L2A: Optical. IT measures Land cover, vegetation health, water quality, urban growthAtmospheric correction applied (reflectance values ready for analysis), We are using 16 bands from 1 to 12 and then four more bands like AOT, WVP.

Landsat-8/L2: Optical + Thermal. Surface reflectance + land surface temperature, Over 10 spectral bands, including thermal, corrected for atmosphere

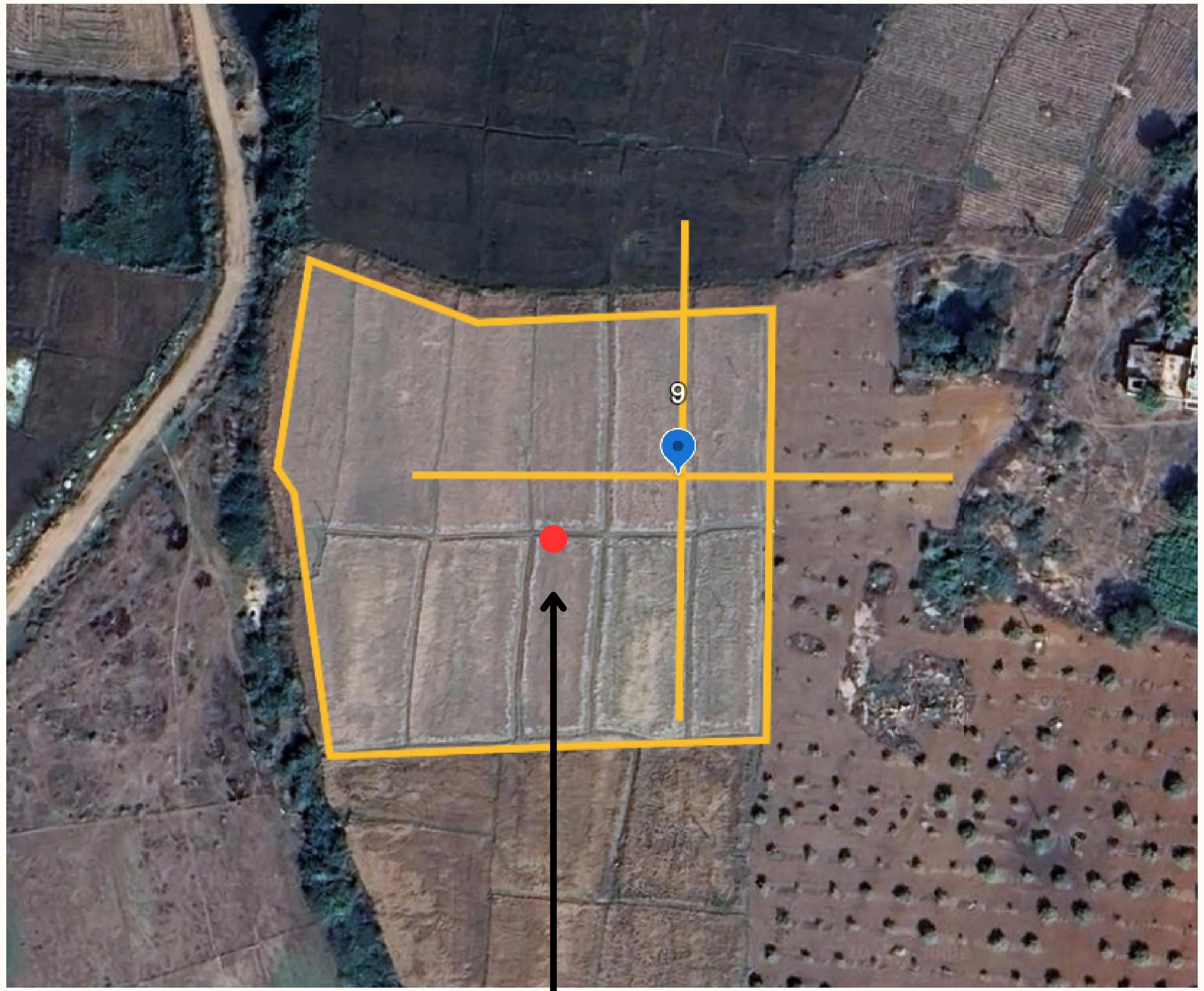
METHODOLOGY (MASKS)

We used Google Earth to prepare our plot masks. We drew polygons in the Google Earth editor and exported the masks as KML which could be further processed into shapefiles or GeoJSON as needed.

Since we had latitudes and longitudes upto 3 decimal places , we follow a two step process,

- First we identified coordinates from our newer dataset and then we dropped pins on each individual coordinate.
- Then we searched a full 50-60m radius to understand where the plots are actually supposed to be.

Let us now discuss the mutually decided rules which were made to ensure consistency while handling ambiguities.

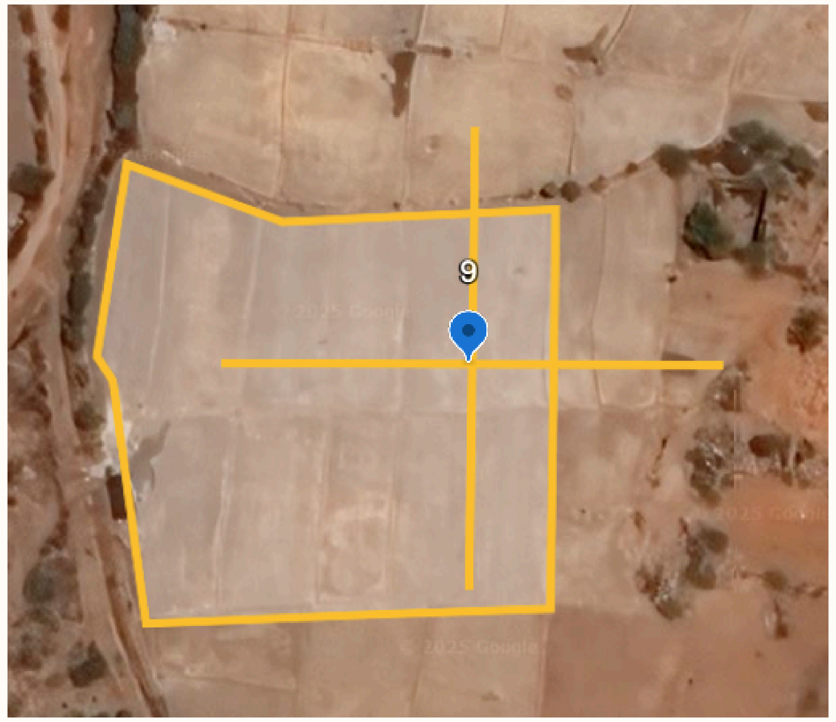


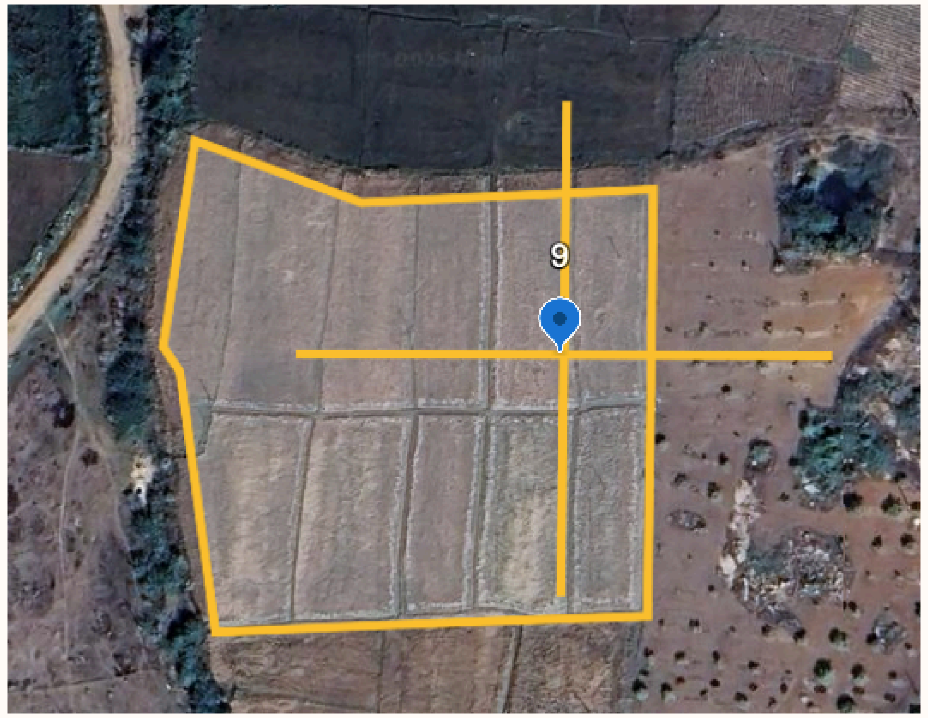
Notice how centroid of annotated plot lies inside square!

ANNOTATION EXPLANATIONS

Use timeline to isolate plots!

- A lot of plots remain barren in some periods.
- We do NOT account multi cropping on the same plot!





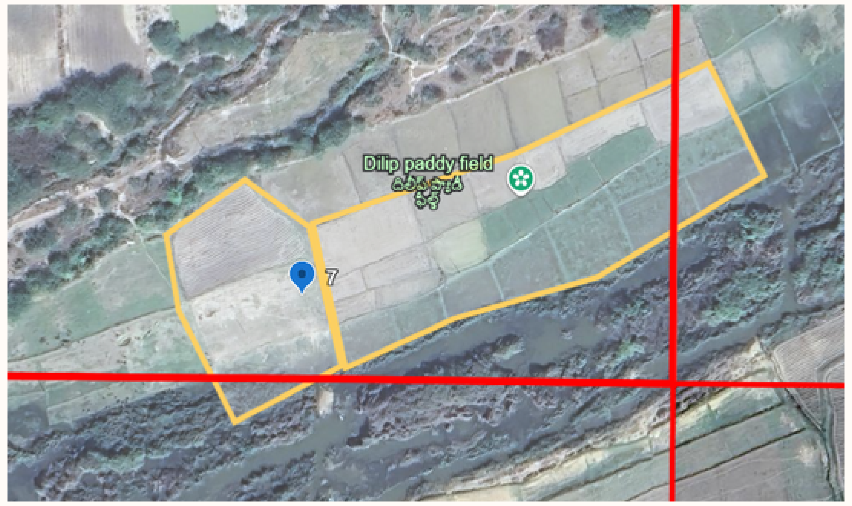
ANNOTATION EXPLANATIONS

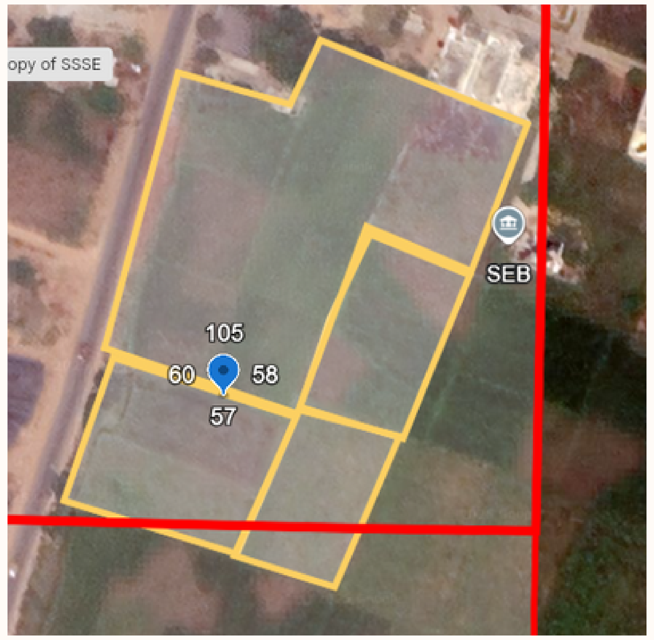
There can be:

- Multiple plots for one point
- Multiple points for one plot
- Multiple points for multiple plots!

Mark them all!

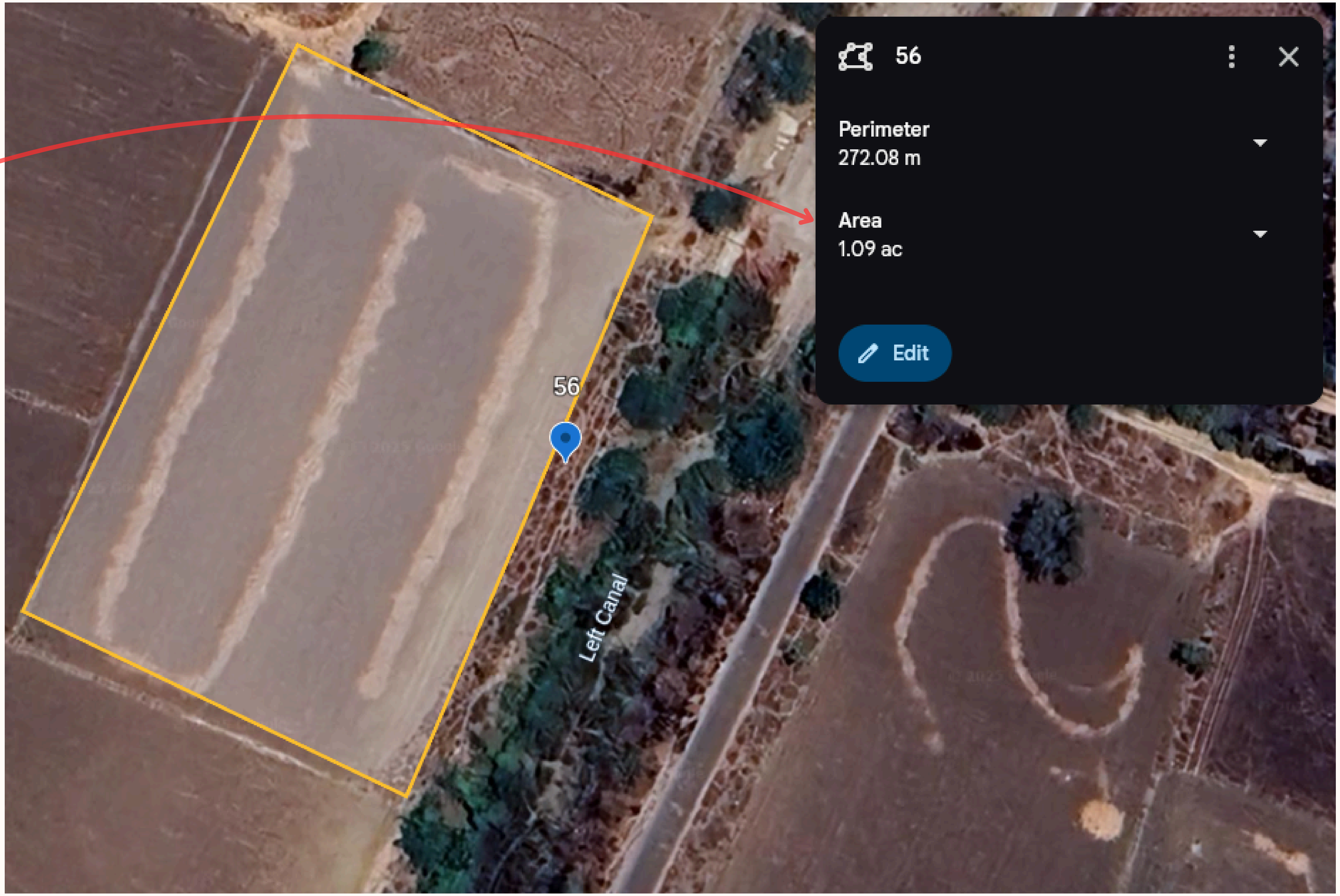
Despite annotations, we only use one-to-one mappings in our current analysis!





ANNOTATION EXPLANATIONS

Plot Area Mismatch!

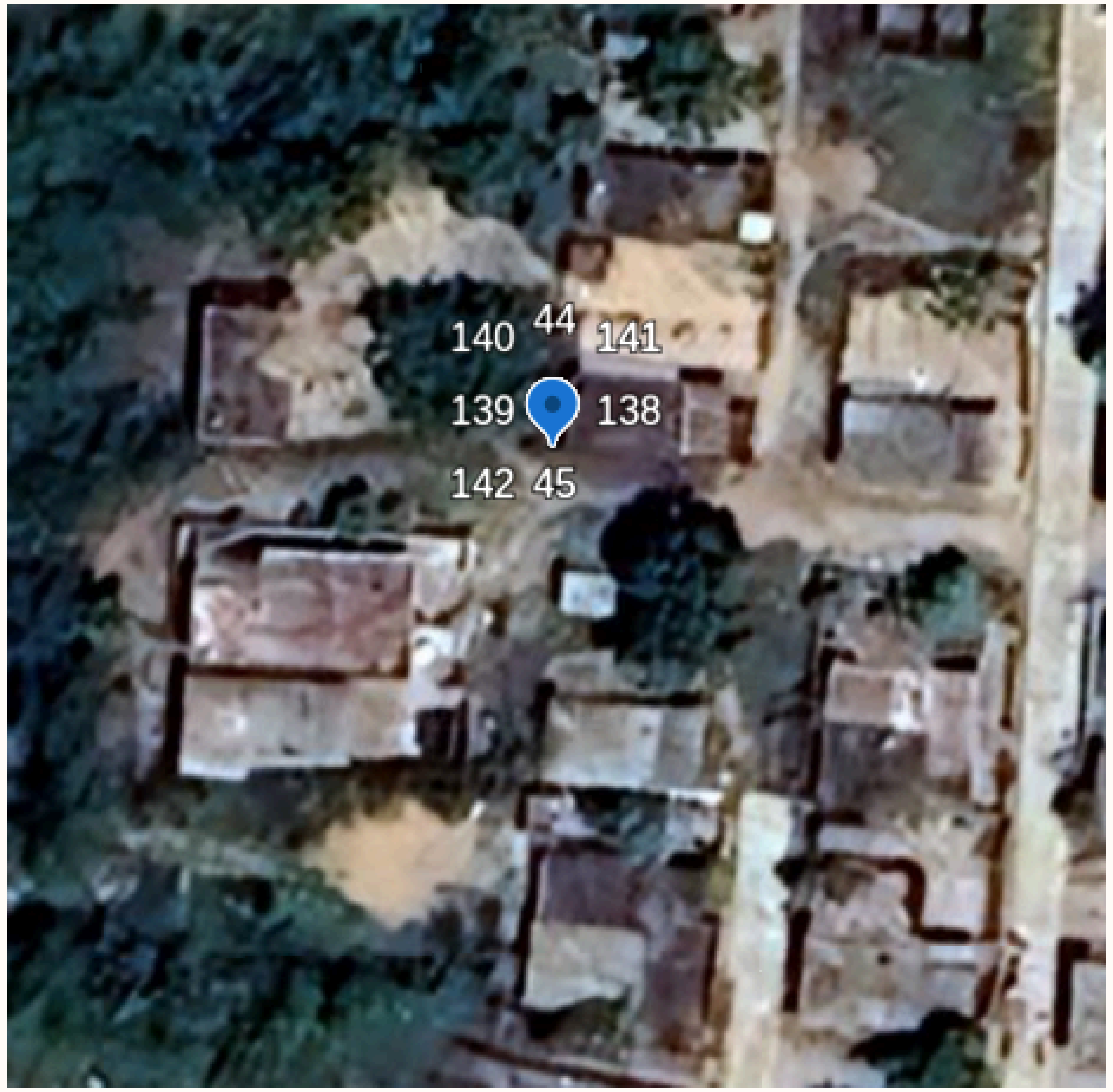


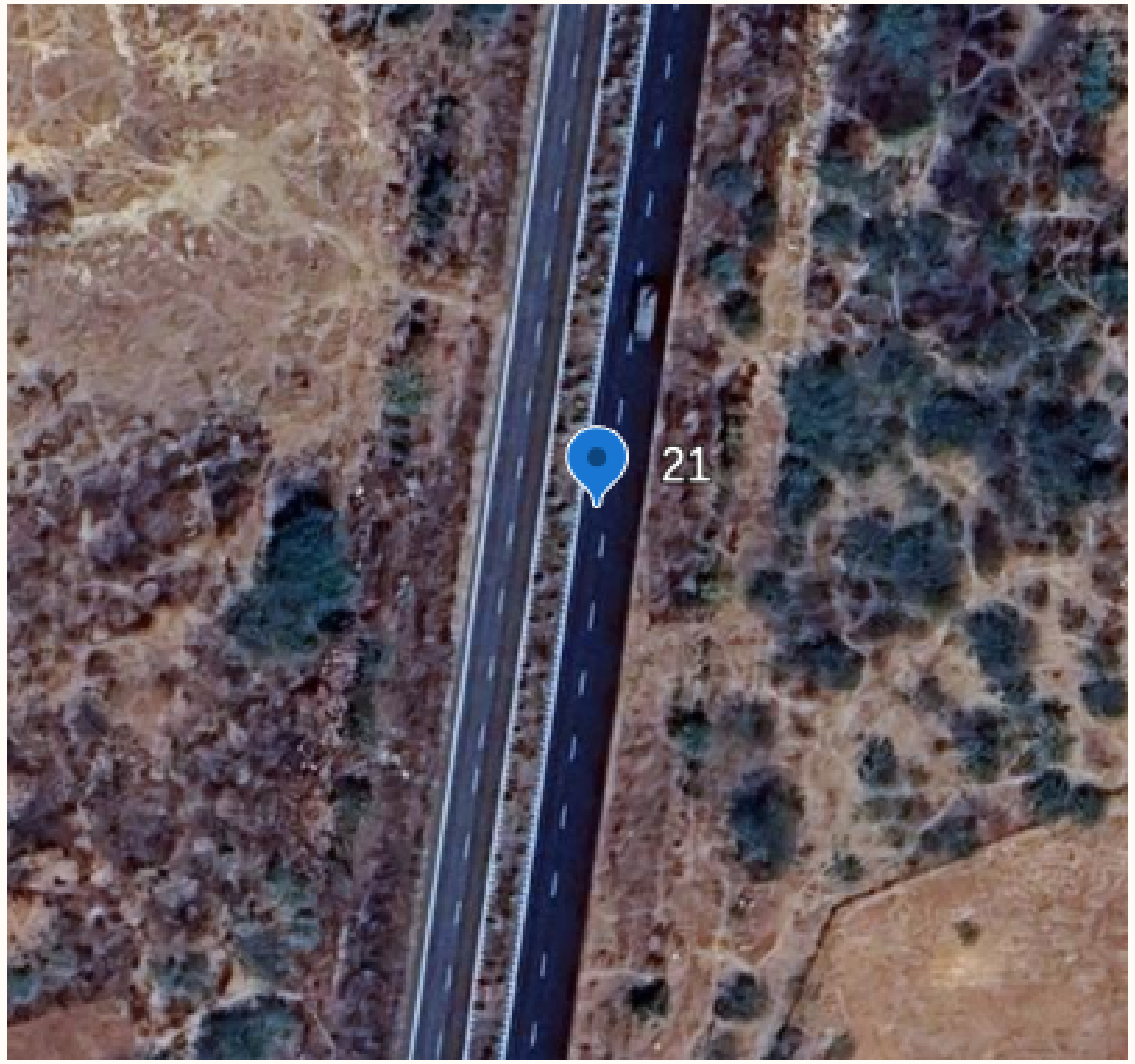
Original dataset said the area is 0.5 acres! None of the candidate plots had an area of 0.5 acres.
This issue was also very common and we decided to ignore it and focus more on the plot locations instead.

ANNOTATION EXPLANATIONS

Bad data points

Some data points had no farm plots at all, or were badlands.





ANNOTATION COUNT

We start off with 164 data points:

```
df = pd.read_csv("andhra_rabi.csv")
len(df)
```

✓ 0.0s

```
164
```

Among which there are only: 73 unique latitude and longitude pairs.

Several data points have the same latitude and longitudes.

For these, we annotated 118 plots for 88 unique data points. There were no suitable plots for the rest.

Then, 42 of these plots have multiple possible data points associated with them.

And, 39 plots are non-unique (multiple candidate plots for a single data point).

Then, we generate a fishnet and find cells which intersect with the plots.

80 such cells exist.

We then remove any masks which have plots of different data points and are left with 56 points.

We then have to remove 3 more points because their growing season was outside of our selected time span (Aug to Dec).

ANNOTATION NOMENCLATURE

- On google earth, save the given latitude and longitude as well as the candidate plots. Name the plot polygon as the index of the corresponding lat-long pair.

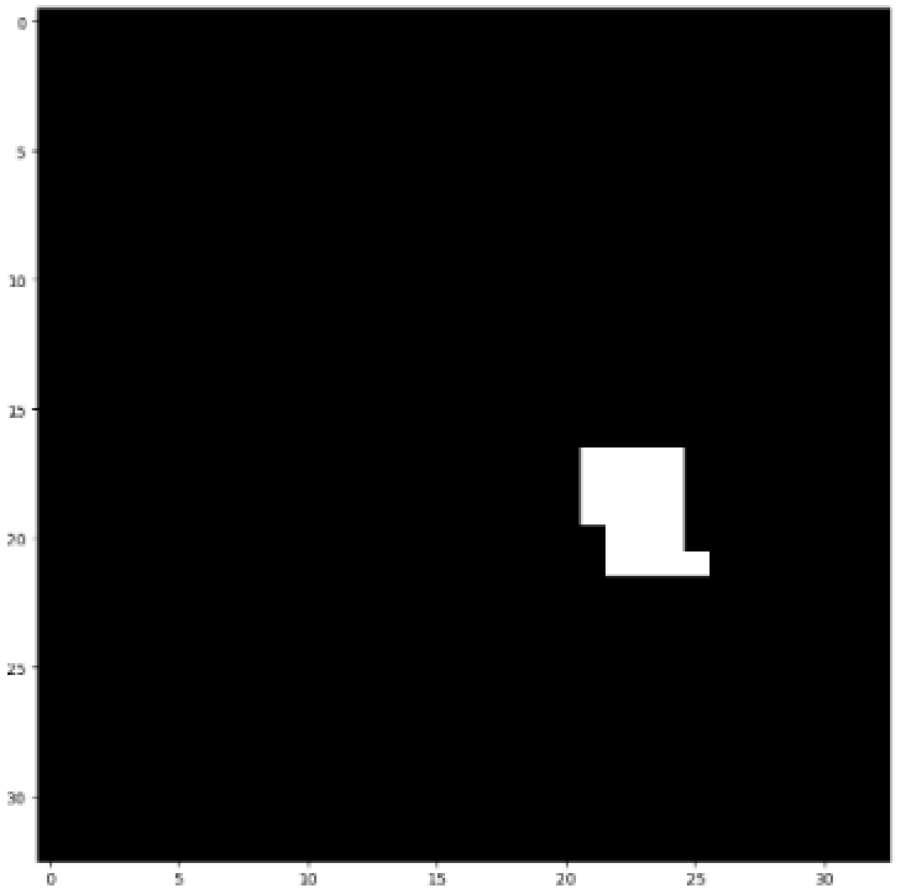
- In case there are multiple corresponding lat-long pairs, include all of them in the name (separated by commas).

- We do not worry about unique ID in plots here since that can be easily taken care of in python.

ANNOTATION TO MASK

Then we divided up our region of interest into grid squares of 330x330m (notice red boxes) to create grid cells which would later be converted into masks that follow the SICKLE format.

Plot mask and Grid Cell

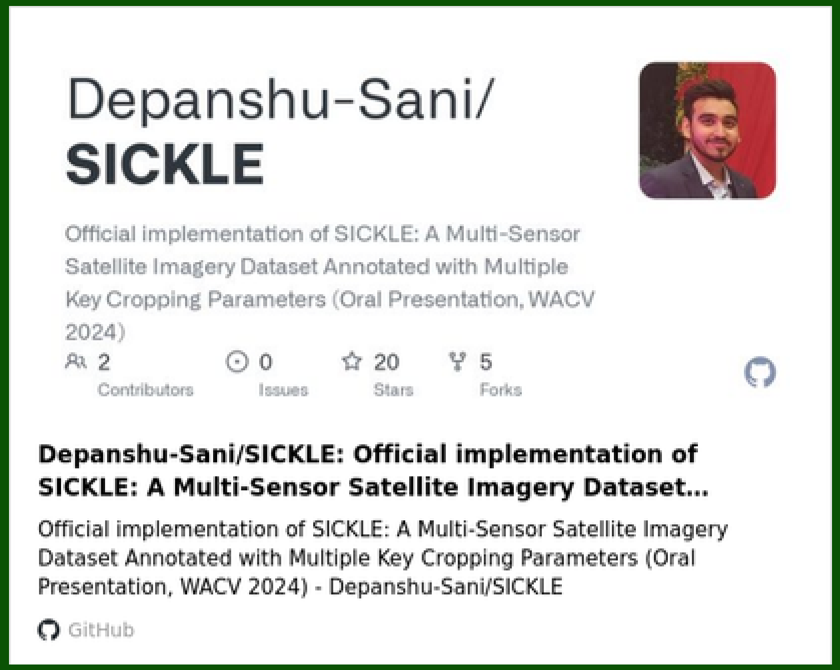


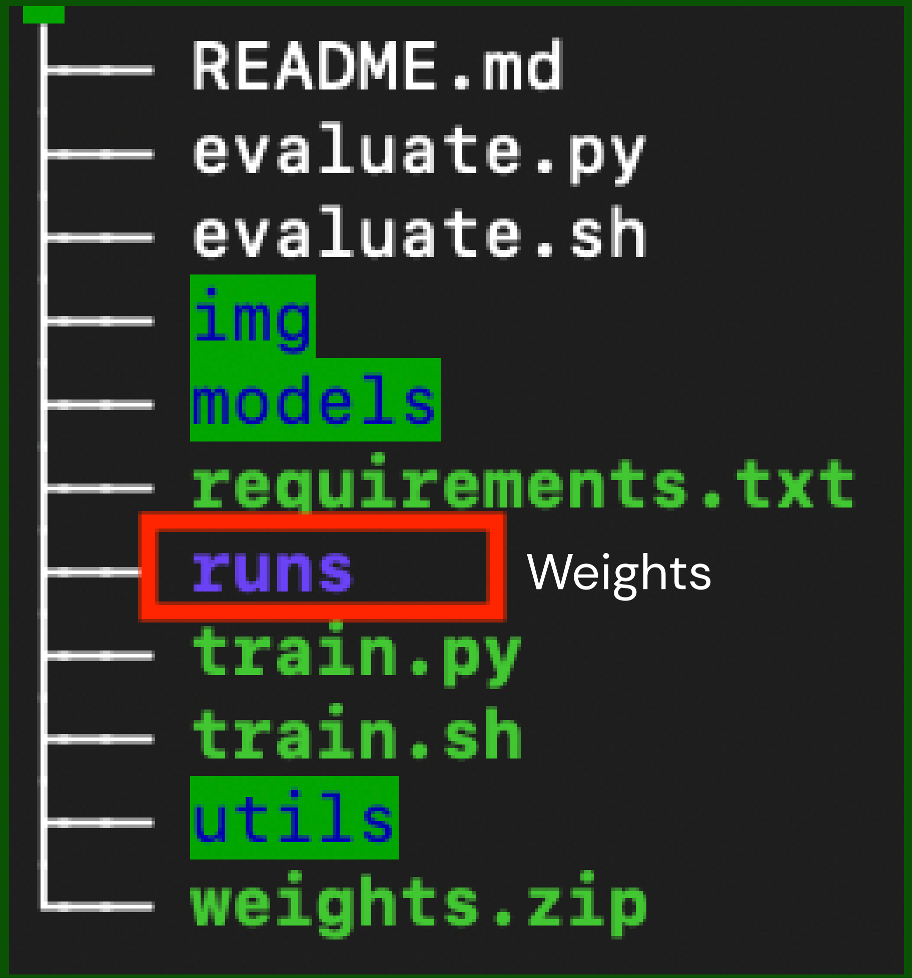
SICKLE Mask

UNDERSTANDING THE GIVEN CODE AND ECOSYSTEM

Inference with pre-trained models

Pre-trained SICKLE weig`hts were made available and we could directly use them to run inference.
The code also provided scripts to do run inference and evaluate metrics.





Repository Structure

UNDERSTANDING THE GIVEN CODE AND ECOSYSTEM

Trained Models

3 models were trained in the SICKLE paper:

1. ConvLSTM
2. UNet 3D
3. UTAE

Tasks

There was one classification task and two regression tasks

1. Crop type prediction
2. Crop phenology prediction
3. Crop yield prediction

We can run the evaluate.sh scripts to get inference.

Example: ./evaluate.sh ../toy_data_new "[S1]" unet3d

EVALUATION METRICS

Classification

1. **F1 Score:** The harmonic mean of precision and recall. It balances the trade-off between false positives and false negatives.
2. **Accuracy:** The proportion of correctly predicted instances among the total predictions.
3. **IoU (Intersection over Union)**: Measures the overlap between predicted and true labels as a ratio of their intersection over their union.

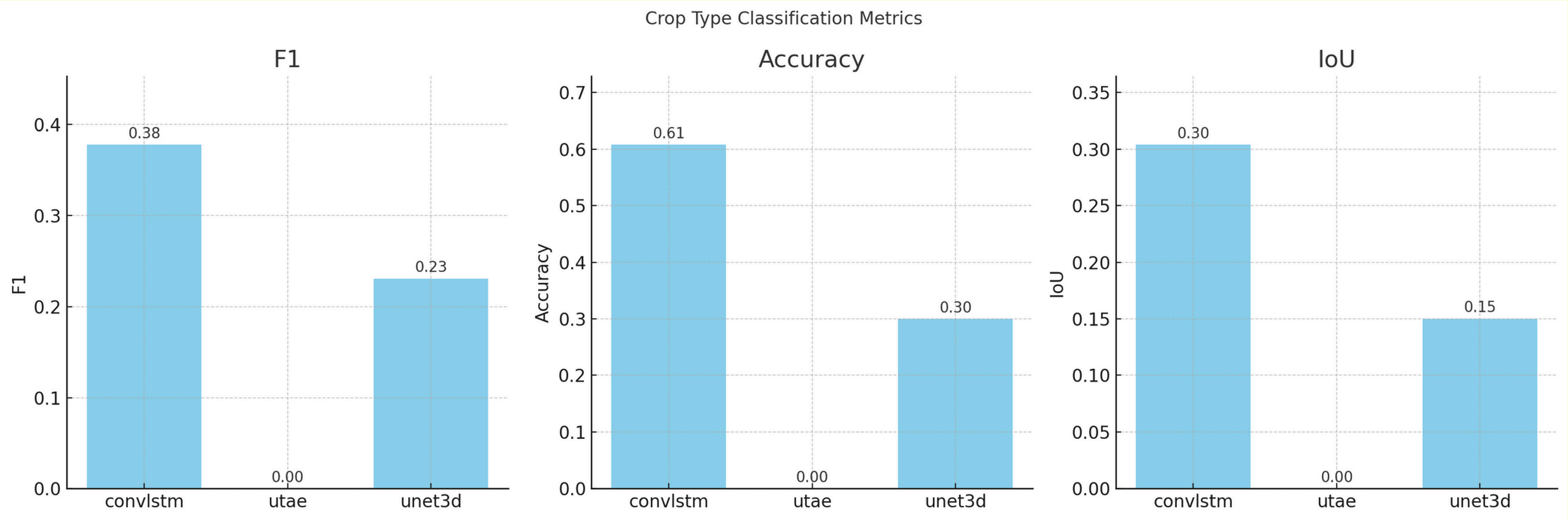
Regression

1. **RMSE (Root Mean Squared Error):** Measures the square root of the average squared differences between predicted and actual values.
2. **MAE (Mean Absolute Error):** The average of absolute differences between predicted and actual values.
3. **MAPE (Mean Absolute Percentage Error):** The average of absolute percentage errors between predicted and actual values.

RESULTS: CROP TYPE

	Overall			Paddy		
Model	F1	Acc	IoU	F1	Acc	IoU
U-TAE	0.7259	0.8028	0.5904	0.5806	0.7670	0.4091
UNet3D	0.9163	0.9474	0.8496	0.8653	0.9489	0.7626
ConvLSTM	0.9163	0.9474	0.8496	0.8653	0.9489	0.7626

SICKLE's Results



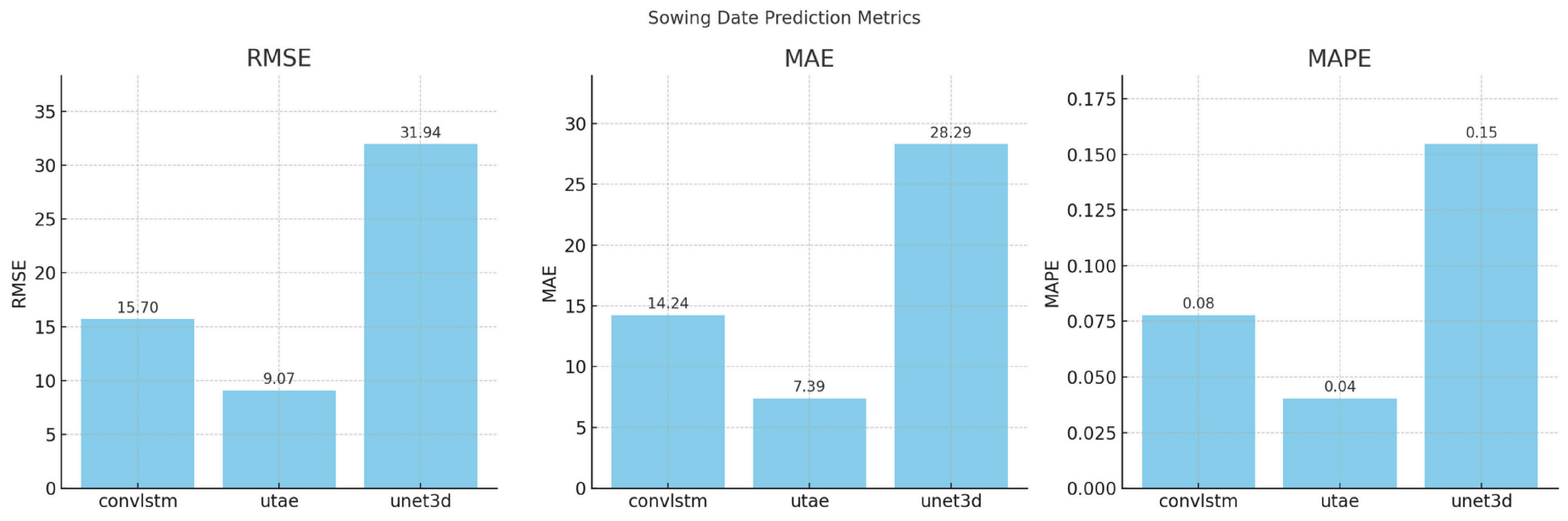
Our Results

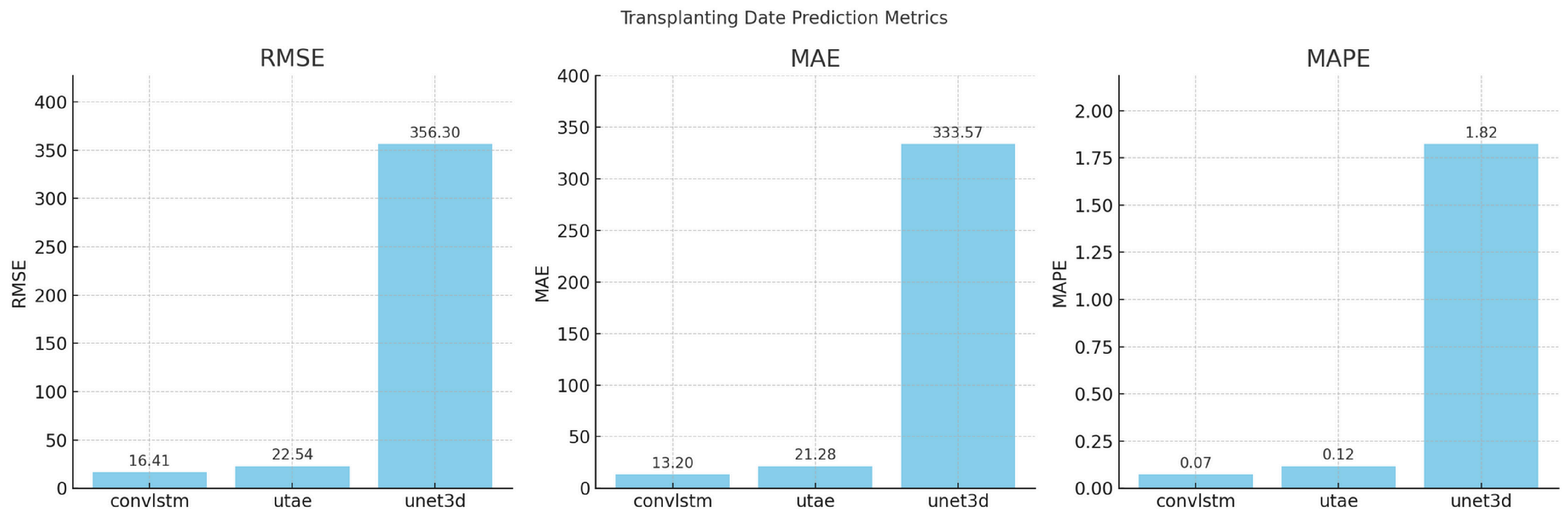
ConvLSTM performs the best here.

Note: UTAE results for this task were not available due to technical issues.

RESULTS: CROP PHENOLOGY

Our Results





SICKLE's Results

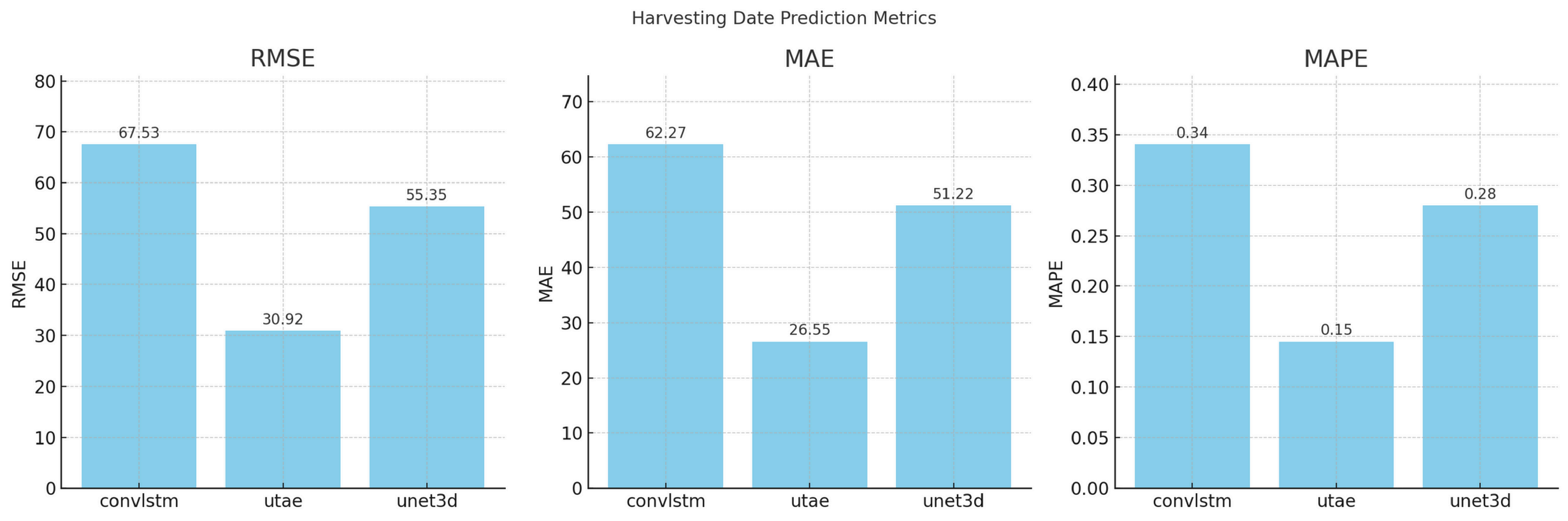
Model	RMSE	MAE	MAPE
U-TAE	8.5669	5.8880	0.0322
UNet3D	3.2568	2.7558	0.0151
ConvLSTM	3.2568	2.7558	0.0151

Model	RMSE	MAE	MAPE
U-TAE	3.3659	2.5914	0.0142
UNet3D	3.3942	2.5186	0.0138
ConvLSTM	3.3942	2.5186	0.0138

RESULTS: CROP PHENOLOGY

Model	RMSE	MAE	MAPE
U-TAE	12.9196	11.6916	0.0639
UNet3D	13.7296	12.3700	0.0676
ConvLSTM	13.7296	12.3700	0.0676

SICKLE's Results

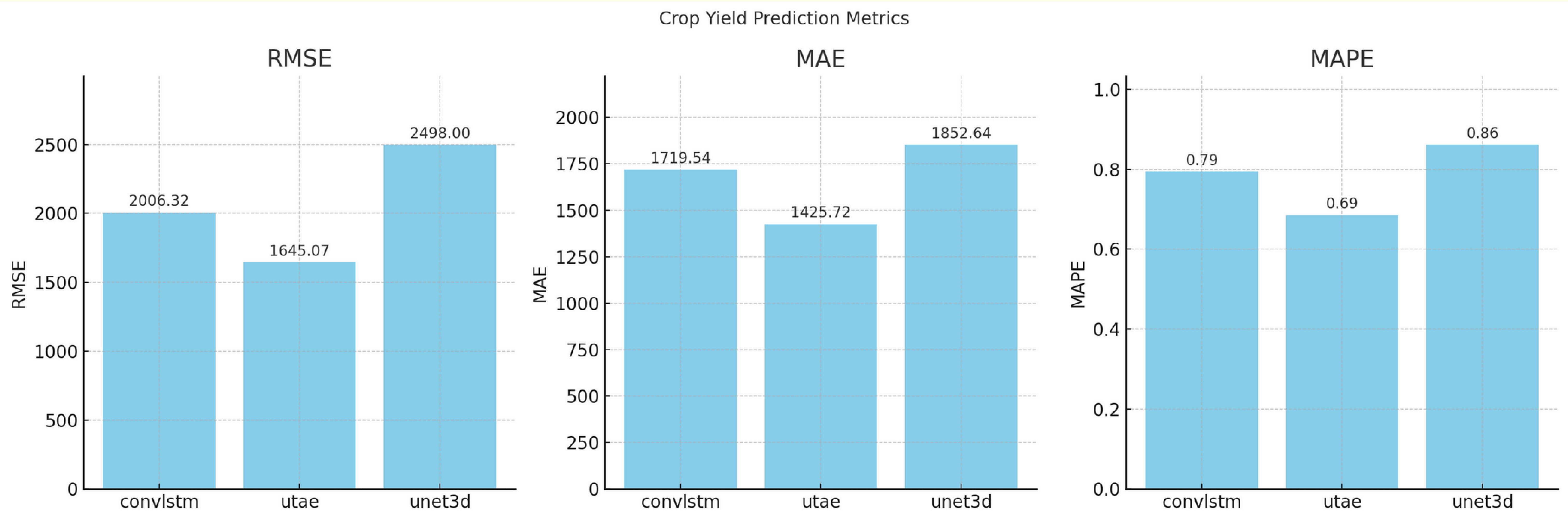


Our Results

RESULTS: CROP YIELD

Model	RMSE	MAE	MAPE
U-TAE	720.7391	562.2070	0.3552
UNet3D	735.4064	561.4650	0.3741
ConvLSTM	735.4064	561.4650	0.3741

SICKLE's Results



Our Results

UTAE seems to perform the best for regression tasks.

CONCLUSIONS

- UTAE outperforms other models across most tasks, especially in temporal predictions (sowing, transplanting, harvesting dates) and crop yield, showing lower RMSE, MAE, and MAPE values.
- ConvLSTM performs well in crop type classification, achieving the highest F1 score (0.378) and accuracy (0.608), but its regression performance is generally weaker than UTAE.
- UNet3D underperforms across tasks, particularly in transplanting date prediction with extremely high errors, indicating poor suitability for temporal regression tasks.
- Model selection should be task-specific: ConvLSTM is preferable for classification (crop type), while UTAE is better suited for regression tasks (dates, yield)

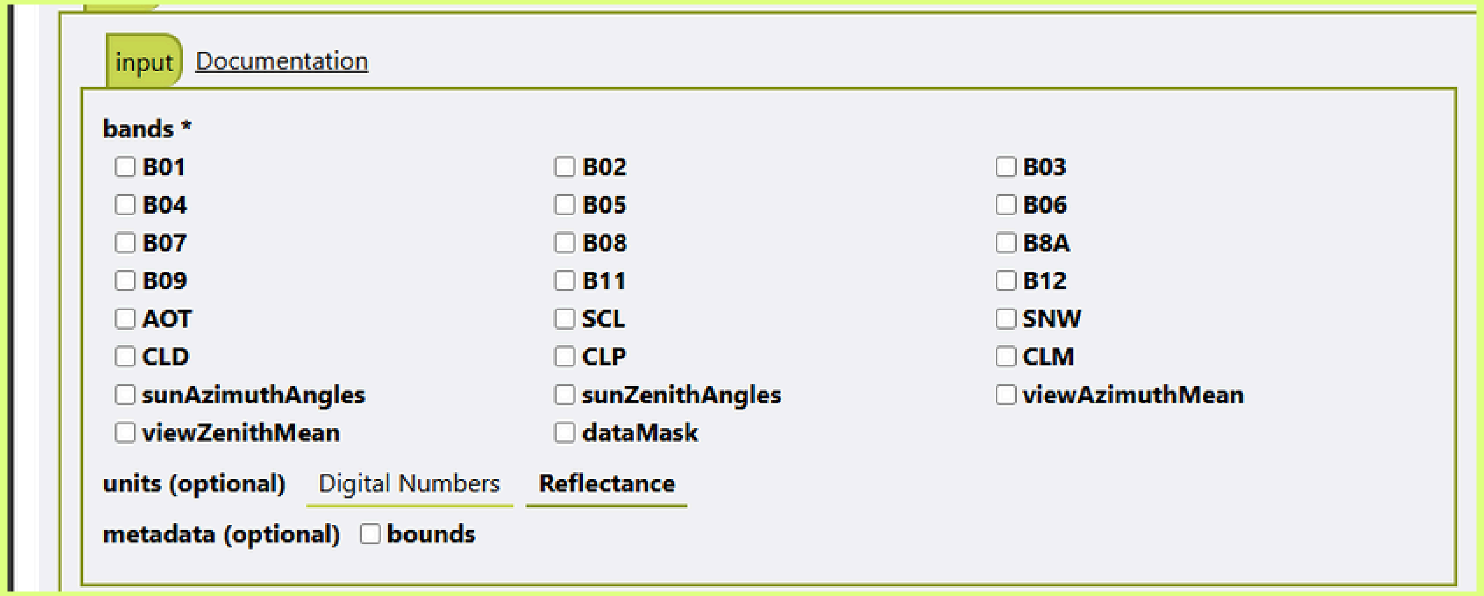
LIMITATIONS AND MITIGATIONS

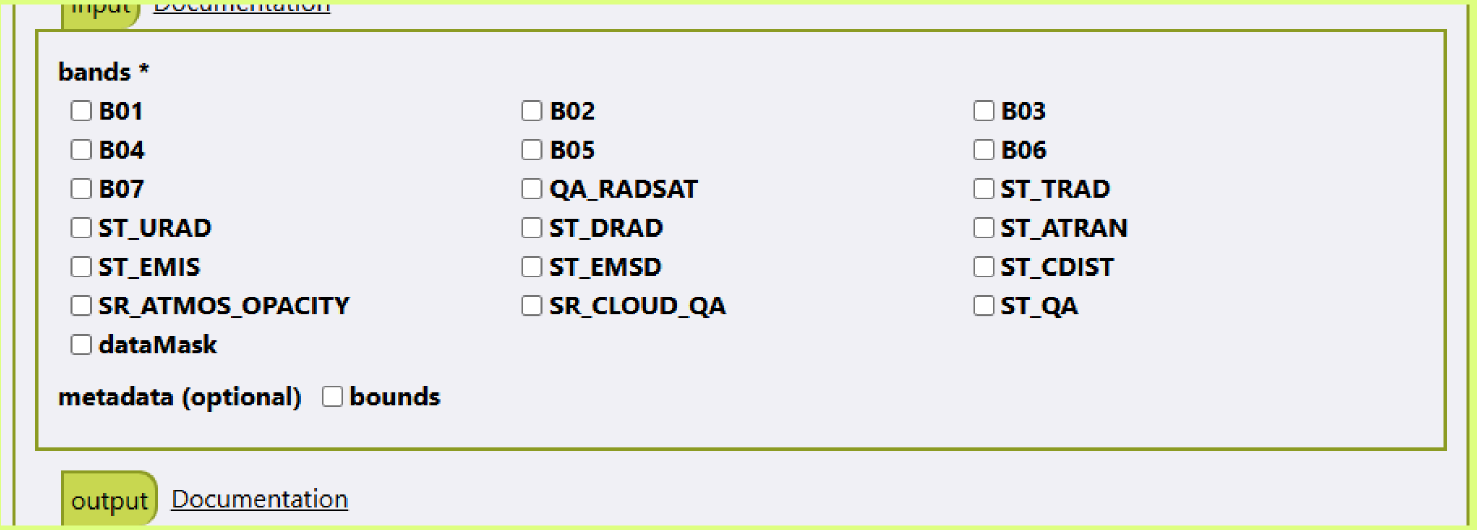
We used the SentinelHub Api to retrieve our plot images. After plotting out the areas, we retrieved various GeoJSON files of different plots, such as bounding boxes, and a satellite image that corresponds to a particular area.

The main problem was with Sentinel-2 data as it was the hardest to recreate. We did not have sufficient information on the number of bands.

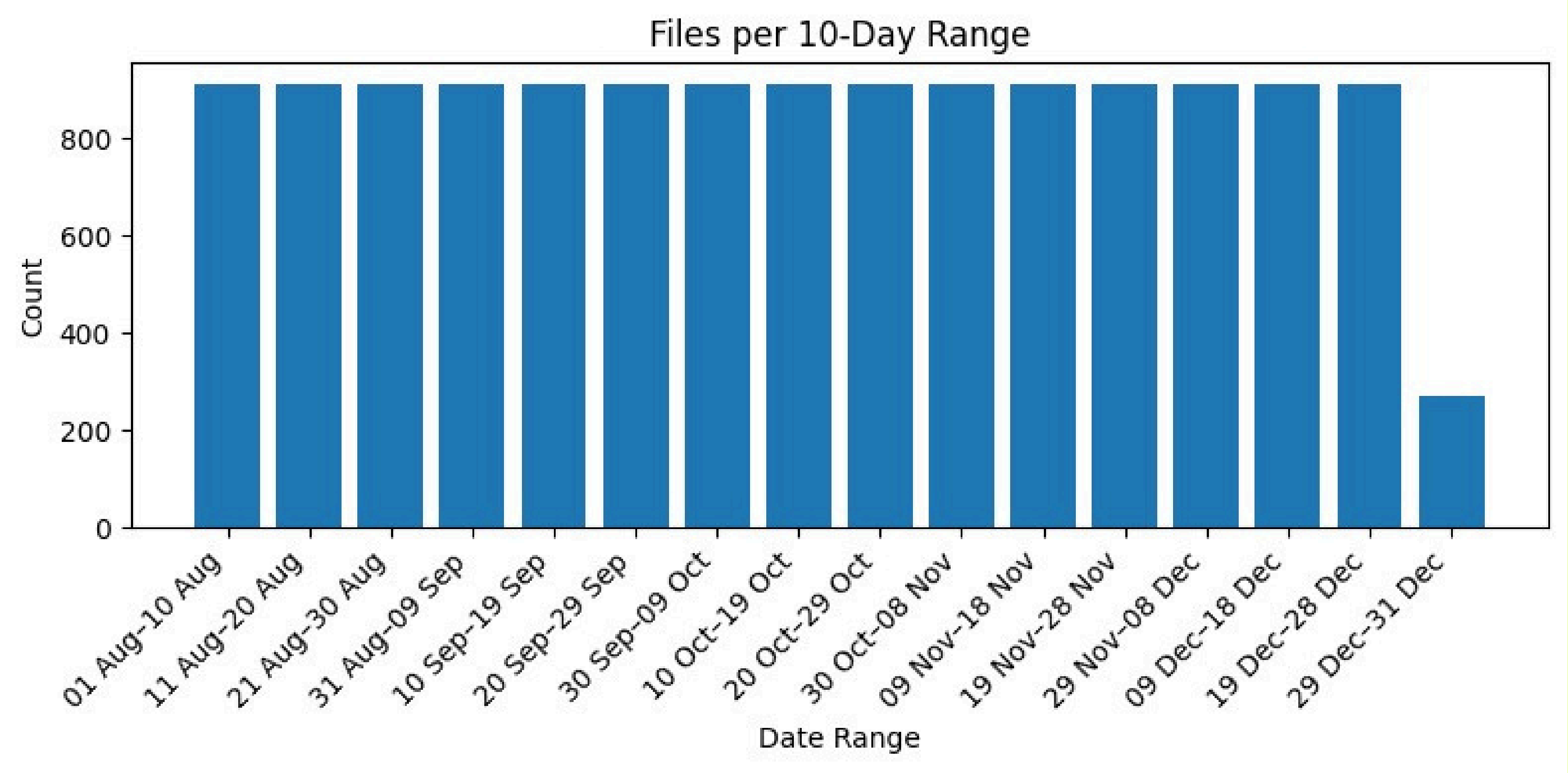
We substituted bands like B08 and B09 for B08A and WPV. Which had the closes resemblance to the real world application, however an option like that exists but it is not downloadable.

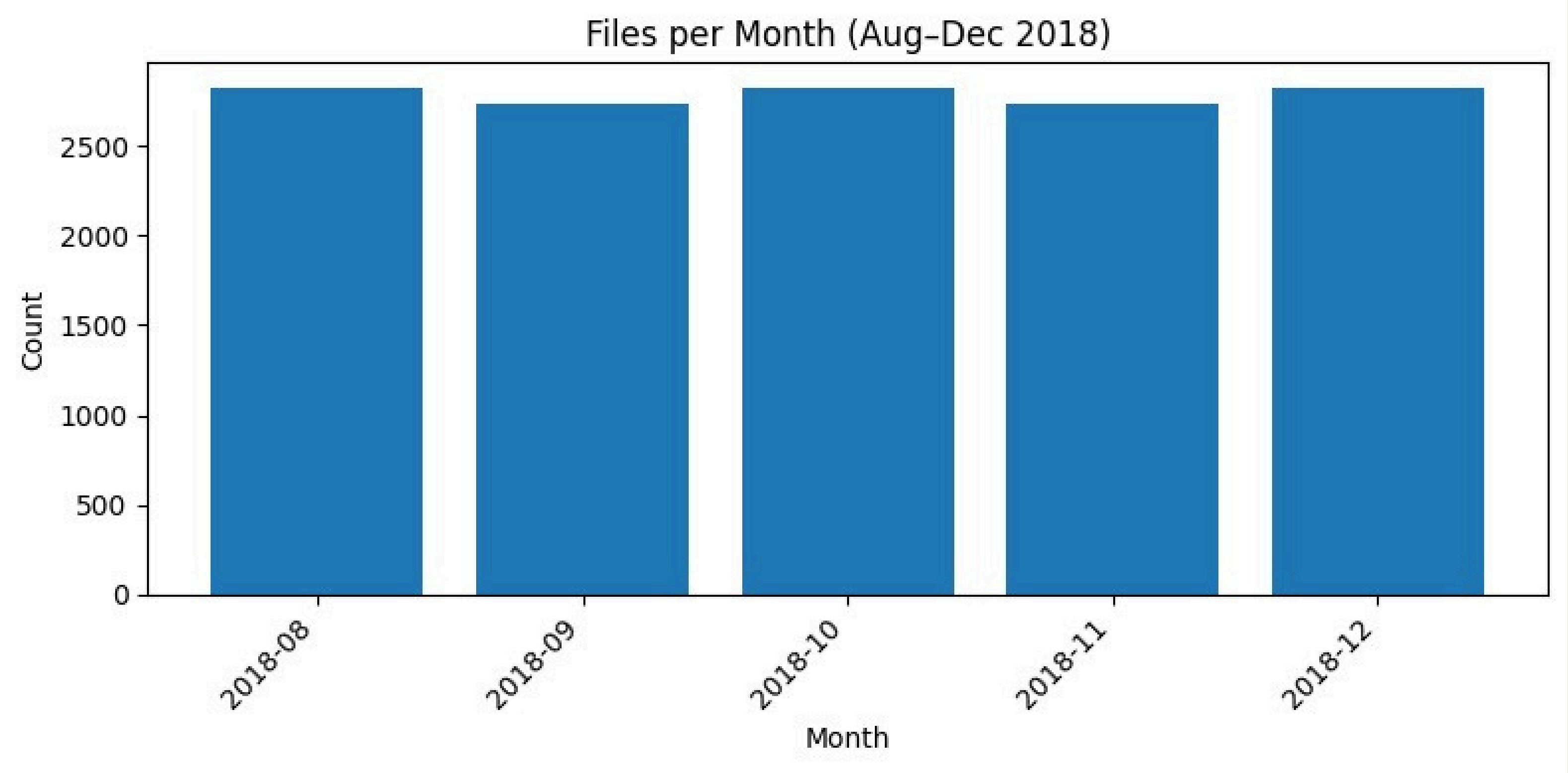
For LANDSAT-8 however, the data bands were not available for download and there was a lot of ambiguity from which source to choose from as they all yeild same data. Also band SL-B10, is not available neither any close substitute.





MISSING DATA VALUATION FOR (S2)





A lot of data for S2 is not available at larger scale as a whole i.e there were dates within our date range for our study(1 Aug 2018 to 31 Dec 2018). But for our little bounding boxes we managed to get sufficient data.

There were some hidden bands which were not mentioned in any written source and we had to brute force them to find out. The bands are AOT, SCL, WPV. The final was a cloud cover band which was assumed to be default for this as there is no other parameter to tweak it further.

LIMITATIONS

- All the challenges in making masks that were mentioned before.
- We used the standard Rabi seasons as our time span instead of the actual season that the plots might have been cultivated in. This can add into the inaccuracy. SICKLE has support for finding inference on actual season as well but we lacked the data for this.
- Lack of data points: we only considered the points which were ambiguous in terms of their respective plot due to time constraints and complexity. This brought us down to just 53 data points.

LIMITATIONS

- **How to handle crop yield?** We had per-acre crop yield values. We also had the plot areas. So, we can calculate the total yield for each plot, right? No! remember the plot area issue we mentioned while annotating the masks. Plot areas are not accurate.
- Two methods:
  - Using per-acre values.
  - Using per-pixel values (dividing the yield by the number of pixels).
- We ended up using the per-pixel method since it performed significantly better than per-acre values but the predictions were still quite inaccurate.

LIMITATIONS

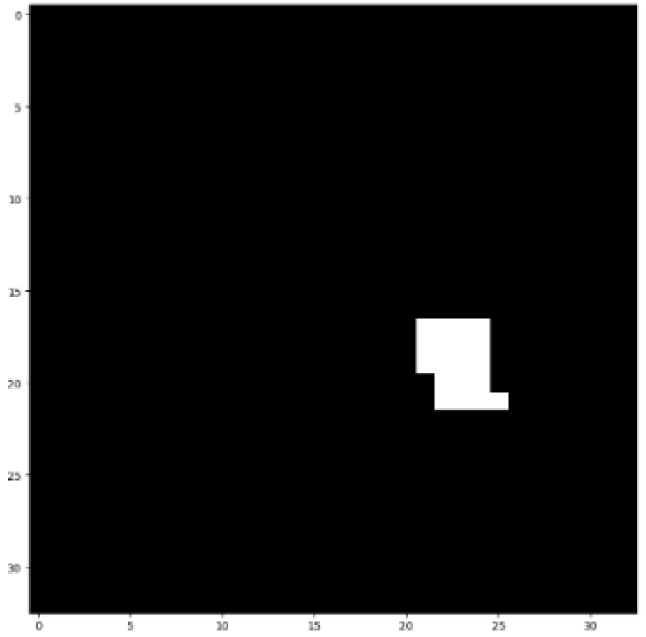
Another issue:

- Re-assigning yield levels from irregular polygons (plots) in the real-world to square-shaped plots introduces a rather non-trivial error.
- Yield levels 96-99.99 and 100.01-103 implies about ~7% error with the truth reference being 100 kg at the pixel level.
- **T**he above error is potentially compounded by edge-effects, I.e. the fact that all 13 (or 14) pixels are unlikely to lie within the plots. This means yield levels assigned in step 1 or 1a are further misrepresented for pixels that partially lie outside the plot.

1.5 acre plots

Total output = 1350 kilograms
Number of pixels in each plot = 13-14 depending on the spatial configuration)
Pixel-yield = 96-103 kg/pixel

Plot mask and Grid Cell



SICKLE Mask

Notice the uneven edges

FUTURE WORK

1. **Getting more data points**: Espically the ambigious points that we decided the drop. We can also consider other seasons as well from the original CSISA dataset.
2. **Modelling the area differences:** We can further model the differences in the actual plot areas and the data to interpret our errors in terms of the difference between the area. We might be able to explain some results through the area difference.
3. **Mask resolution**: We can perhaps try higher resolution masks to reduce the edge effect. This might improve the crop yield performance or we can still test the hypothesis that the uneven edges cause problems.
4. **Training SICKLE:** If we are able to gather enough data, we can try to retrain SICKLE on Andhra Pradesh region and see the differences between the original SICKLE predictions and the retrained-SICKLE predictions. This can help us rule out whether the error occurs due to the region differences or other issues.
5. **S2, L8, Fusion Model:** We were unable to get S2 and L8 data. Once available, we can try the fusion model as well.


THANK YOU