Combining Satellite and Weather Data for Crop Type Mapping: An Inverse Modelling Approach

Praveen Ravirathinam[†]   Rahul Ghosh[†]   Ankush Khandelwal[†]   Xiaowei Jia[*]

David Mulla[†]   Vipin Kumar[†]

Abstract

Accurate and timely crop mapping is essential for yield estimation, insurance claims, and conservation efforts. Over the years, many successful machine learning models for crop mapping have been developed that use just the multi-spectral imagery from satellites to predict crop type over the area of interest. However, these traditional methods do not account for the physical processes that govern crop growth. At a high level, crop growth can be envisioned as physical parameters, such as weather and soil type, acting upon the plant leading to crop growth which can be observed via satellites. In this paper, we propose Weather-based Spatio-Temporal segmentation network with ATTention (WSTATT), a deep learning model that leverages this understanding of crop growth by formulating it as an inverse model that combines weather (Daymet) and satellite imagery (Sentinel-2) to generate accurate crop maps. We show that our approach provides significant improvements over existing algorithms that solely rely on spectral imagery by comparing segmentation maps and F1 classification scores. Furthermore, effective use of attention in WSTATT architecture enables detection of crop types earlier in the season (up to 5 months in advance), which is very useful for improving food supply projections. We finally discuss the impact of weather by correlating our results with crop phenology to show that WSTATT is able to capture physical properties of crop growth.

Keywords- Remote Sensing, Spatiotemporal data, Crop mapping, Inverse Modelling, Multimodal data

1 Introduction

With increased need for food due to rapidly growing population and erratic weather patterns, accurate crop monitoring is essential for forecasting food supply and proper management of resources[25]. A crucial aspect of crop monitoring is pixel-wise crop mapping, which assigns a crop category label to each pixel in a specified region. Accurate and timely crop maps can facilitate land

---
[*]University of Pittsburgh. {XIAOWEI}@pitt.edu

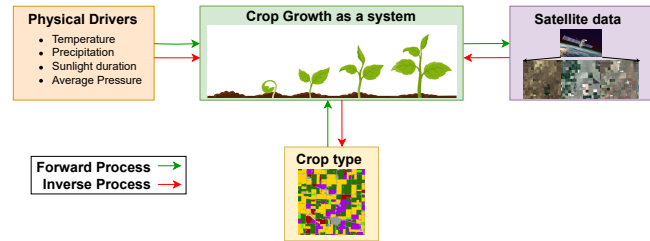[†]University of Minnesota. {pravirat, ghosh128, khand035, mulla003, kumar001}@umn.edu

Figure 1: A diagrammatic representation of the physics point of view understanding of crop growth

use planning, yield estimation, pest management, and the evaluation of sustainable management practices and conservation efforts[2]. Various data-driven approaches, ranging from simple regression-based models [19, 27] to complex deep-learning models[31, 20, 33], have been proposed to build crop maps. A vast majority of these approaches rely solely on satellite imagery sources to generate crop maps for their study region [10, 34, 2].

However, from a physics point of view, crop growth signal observed by satellite images is a result of complex interplay between type of crop, weather patterns, soil, management practices, etc. This view is also referred to as forward modeling of crop growth (shown as green arrows in Figure 1). Since, weather is a major source of heterogeneity (same crop can grow differently due to variation in weather and same weather pattern impact crops differently), we aim to leverage this forward modeling strategy by including weather data into deep learning model for the task of crop type mapping. Such an approach has the potential to offer new insights into understanding and predicting crop dynamics and improve generalizability over space and time. In physics-based models (that simulate the forward modeling of crop growth), crop-type is often considered as one of the initial conditions (or simulation parameters, knobs, etc) because the response of a system to weather drivers is conditioned on these parameters. Such a problem setting occurs in many other applications. Consider the example of predicting stream flow of a watershed, where a similar weather pattern can lead to different streamflow depending on the watershed characteristics (such as forest cover, soil type, terrain,etc.) [12].

Since our goal is to estimate the crop-type label, we cannot use this forward model formulation directly. Specifically, in our case, weather drivers are available as well as a proxy of crop growth (observed via satellite imagery) but the key initial condition, i.e., crop type is missing. To address this issue, we propose to use an inverse modeling [12] formulation where hidden relationships between weather and crop-growth are learned to infer the initial condition (crop-type) (red arrows in Figure 1). In particular, we present WSTATT (Weather-based Spatio-Temporal segmentation networks with ATTention), a deep learning model that combines the spatio-temporal satellite and weather data with attention to give accurate pixel-wise segmentation maps for a given region. The key idea in WSTATT is to create embeddings that capture inherent relationships between weather and crops to improve classification performance. We show the efficacy of WSTATT for crop mapping task over the state-of-the-art model, STATT [10], that relies only on satellite imagery and was shown to be better than various spatiotemporal methods such as 3DCNN[17], ConvLSTM[37], and CALD[18] for crop mapping. Specifically, we show that WSTATT not only performs well on estimating crop labels, but also it can assign these labels without seeing data for the full year. This early prediction capability enables timely intervention, yield prediction, pest management, and risk mitigation. We further show that our attention module effectively combines weather and satellite imagery to captures key discriminative periods among different crops. Finally, we discuss the impact of using weather data in prediction for each crop class by correlating the results to their phenologies.

Our contributions can be summarized as follows:

- We develop an inverse modeling-based deep learning approach that uses both spatiotemporal weather and satellite data for pixel-wise crop type mapping.

- We show that our proposed approach outperforms the current state-of-the-art on both year-end prediction and early-prediction for crop mapping.

- We provide visual analysis of attention weights to show the influence of the inverse modeling approach in effectively identifying discriminative timeframes.

- We qualitatively analyse the relationship between model performance and crop phenology to show the model's ability to learn meaningful patterns.

- We release the code [1]1 used in this work and full length paper to promote reproducibility.

---
[1]Code and Full Paper Link

2 Related Work

2.1 Land Cover Mapping Various land cover mapping methods have been developed that use satellite imagery as their main input, ranging from simple regression techniques [19, 27] to more complex deep learning approaches [31, 20, 33]. Many of these methods involve using convolutional networks to extract spatial features [15, 22], while others utilize spatiotemporal methods to take advantage of the spatiotemporal nature of satellite data [39, 34, 10]. These methods often involve combining recurrent networks [14, 29] with traditional convolution networks, and some even incorporate data from multiple satellites to address issues such as clouds and missing data[3]. Using these methods, many land cover datasets have been created, covering large spatial extents[13, 38, 36, 35, 21]. One notable example is CALCROP21 [11], a large-scale dataset that provides pixel-wise information on both major and minor crops in California. Recently [28] proposed a recurrent neural network based method for early prediction for crop classification using two output heads, one for each classification and stopping probability. They were able to show that their model was able to provide class labels sooner, and with less data. However, their test data was from the same year as that of training, which leads to the question of whether their probability head would carry over to a future year, as the spectral signatures, frequency of data and management practices would all change. For early prediction in years different from the training year,[6] suggested to use target domain data in an adversarial fashion to adapt their model and help in target domain classification. In our work, we show that inclusion of weather data inherently provides adaptation capability.

2.2 Weather based Remote Sensing Crop growth is affected by weather, but its application in crop mapping is limited. Previous studies used weather data to correct their classification approaches [9] while training if the crop growth was inconsistent with meteorological signals. More recent research [7] incorporated macronutrient information and weather data to predict suitable crop types using an SVM-based method. Self-organizing map-based methods were also used to forecast suitable crop growth and weather [24]. In a recent study [26], a combination of weather data, field information, and satellite data was used to track the growth stages of maize and provide decision-making support for maize fields. However, this study assumed knowledge of the specific crop type, which are not available in our setting. Another study such as [1] briefly experimented with incorporating weather and satellite data in field crop type mapping. Although this study also used deep learning with satellite and weather data, it has several

key differences from our approach. The authors propose selectively dropping weather time steps to ensure compatibility with their deep learning approach, whereas WSTATT uses recurrent networks to preserve temporal information across all timestamps. Second, this method only incorporates temporal relationships (and ignores spatial relationships) unlike WSTATT, even though it has been shown that spatio-temporal methods perform well compared to purely temporal approaches [10]. Third, their method also does not allow for early prediction, due to the lack of variability in temporal input sequence length. Finally, they could not assess the impact of weather on their experiments, claiming that their test set was not rich enough. Weather data has been commonly used in crop-related tasks such as selecting suitable crops for maximizing crop yield [16], forecasting crop yield [30, 23], and assessing crop yield [5, 32]. This usage shows that weather data has promising potential for crop-related tasks.

3 Method

3.1 Problem Setting In this paper, we consider crop mapping as a semantic segmentation task, where we aim to assign a particular class label to each pixel of an image dataset. During training, we have the following data sources:

- Satellite image time-series $S = [S^1, \ldots, S^{Ts}]$, where each $S^t \in \mathbb{R}^{L_s \times B_s \times C_s}$ is a satellite image of size (L_s, B_s) at time t with C_s multi-spectral channels.

- Weather image time-series $W = [W^1, \ldots, W^{Tw}]$, where each $W^t \in \mathbb{R}^{L_w \times B_w \times C_w}$ is a weather data image of size (L_w, B_w) at time t with C_w multi-spectral channels. Note that the temporal frequency of the satellite image time series and weather data can be different.

- Labels $Y \in \{0,1\}^{L \times B \times V}$ in one-hot representation, where V is the number of classes. The objective is to predict one label for each pixel given the time series of satellite and weather data over a year.

3.2 Deep Learning model Our deep learning model is inspired by Spatio-Temporal segmentation networks with ATTention (STATT) [10], a recent state-of-the-art segmentation model for land cover mapping. Like STATT, our model uses an encoder-decoder setup with temporal attention to predict a pixel-wise segmentation map, However, the key difference is our formulation of crop mapping task as an inverse problem, using weather data as an additional input. A diagrammatic representation of our proposed deep learning architecture, WSTATT, can be seen in Figure 2. The architecture has two separate encoders: one for the

satellite data $E_s(\cdot;\theta_{Es})$ and one for the weather data $E_q(\cdot;\theta_{Eq})$. This is followed by an attention module $A(\cdot;\theta_A)$, which uses both encoders' outputs. Finally, the decoder $D(\cdot;\theta_D)$ is used to create the maps.

The satellite data encoder $E_s(\cdot;\theta_{Es})$ consists of convolutional and LSTM layers to capture spatial context and temporal dynamics, respectively. The convolutional layers encode each satellite image S^t to Z^t, and create a time series of spatial embeddings $Z = [Z^1,..,Z^{Ts}]$ The parameters in the convolution layers are shared across multiple images in the input time series. To capture the temporal dependencies within the spatial embeddings, this series is passed into a Bidirectional LSTM to get the hidden states $H_{ij} = [H_{ij}^1,..,H_{ij}^T]$, where T denotes the number of timestamps. The LSTM is governed by a set of equations (please refer to full length paper) which use the previous hidden state H_{ij}^{t-1} and cell state C_{ij}^{t-1} to generate the current hidden state H_{ij}^t.

Each crop class has a different growing pattern across the entire year. However, any two crop classes are generally most distinguishable only during a certain time window, which we refer to as the discriminative period. To effectively capture this period, examining the forward and backward directions across all available timestamps is necessary, leading to the use of Bidirectional LSTM in our method. The Bidirectional LSTM produces two hidden states for each timestamp, which we concatenate to create the final series of hidden representations $H_S = [H_S^1, \ldots, H_S^T]$.

Similarly, the weather data encoder $.E_q(\cdot;\theta_{Eq})$, consists of a Bidirectional LSTM to encode the weather data into its respective series of hidden states $H_W = [H_W^1, \ldots, H_W^T]$. Compared with satellite images, weather data are often collected at a higher temporal frequency (e.g., daily) but at a coarser spatial resolution. As a result, in many cases, a training patch will have only a single weather pixel within the patch bounds. Hence, we do not use convolutional layers for encoding the weather data, and use only a Bidirectional LSTM encoder resulting in a hidden embedding for each timestamp of the weather data. As a result, the length of the weather embeddings list (H_W) is equal to the number of timestamps of weather data present. Since weather data and satellite data are available at different temporal frequency, the vectors H_W and H_S will be of unequal lengths. To address this issue, we select weather data embeddings at the same time intervals as the satellite data's frequency. Specifically, if the satellite data is available every 15 days and weather data is available daily, we would take every 15th weather data embedding to match the frequency of satellite data. Since each hidden state is created using previous hidden states, we
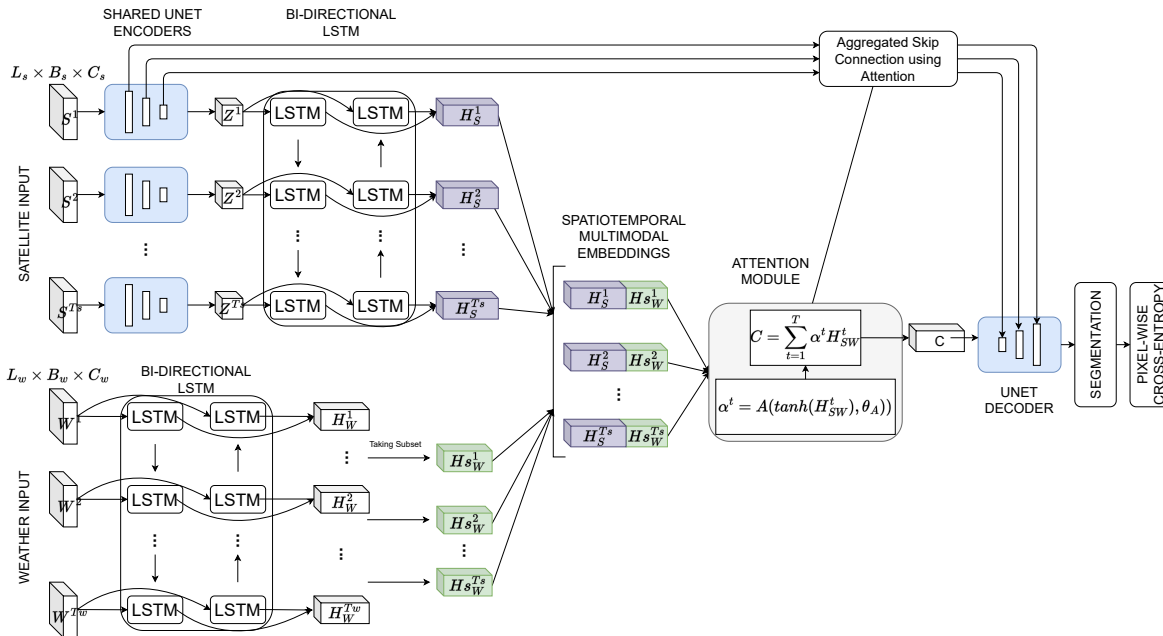
Figure 2: A diagrammatic representation of our proposed deep learning architecture WSTATT (Weather based Spatio-Temporal segmentation networks with ATTention)

still use information from all weather data even though we take a subset of embeddings. So even if, for example, the 15th day has noisy weather readings, since the 15th weather embedding is created using this noisy reading and all previous weather readings, it will still contain relevant information. We also resorted to equally spaced selection subsetting as opposed to averaging-based subsetting because we did not want dynamic changes in weather to be lost in the averaging. After this equally spaced selection of embeddings, we are left with a subset of weather data embeddings $Hs_W = [Hs_W^1, \ldots, Hs_W^{Ts}]$, the same length as H_S. Now, to match the spatial dimensions of H_S, each entry in Hs_W is resampled spatially using the same value (as there is only one embedding value for each H × W of satellite data. Futhermore, we concatenate H_S and Hs_W along the channel dimension to create $H_{SW} = [H_{SW}^1, \ldots, H_{SW}^{Ts}]$. This series represents the spatiotemporal multimodal embeddings of both the satellite images and the weather data, thus enabling the inverse modeling approach by forcing all subsequent components in the method to use both modalities jointly.

This multimodal embedding series is then passed onto the attention network $A(\cdot; \theta_A)$, which assigns a weight for each timestamp dynamically. The attention weight for each timestamp ranges in (0,1) and sums up to 1 over all the timestamps. The weight represents the importance of data at each timestamp towards the final goal of crop mapping. We used a single-layer feed-forward network as the attention layer for our

implementation. This layer helps the model focus on which timestamps are more discriminative for any given class and can also help reducing the effect of issues such as cloud cover blockage or missing data. The series H_{SW} is aggregated temporally using these attention weights (α^t) to form the final embedding C_{SW}.

This attention-aggregated embedding is then sent into the decoder. The decoder $D(\cdot; \theta_D)$ is a set of convolutional layers, similar to the UNET deconvolution approach. Since the input to the decoder is a multimodal attention aggregated series, the decoder is forced to learn the relationship between these embeddings and the final crop map, which is the idea behind the inverse modeling paradigm. Similar to STATT, we also use aggregated skip connections using attention, using the attention weights (α^t) at every step of the decoding. Finally, we use a linear layer followed by softmax to get the pixel-wise class probabilities. The model can be trained using pixel-wise cross-entropy loss.

4 Dataset and Experimental Setup

4.1 Study Region We selected the crop belt in Central Valley of California as our study region (100km x 100km) due to the richness in crops grown and its weather variability. Specifically, we chose the regions bounded by the T11SKA Sentinel-2 Tile, an area with over 30 crop classes. A visualization of the crop classes in this region can be seen in Figure 3. These class labels are taken from the USDA Crop Data Layer (also done by other works such as CALCROP21 [11]).
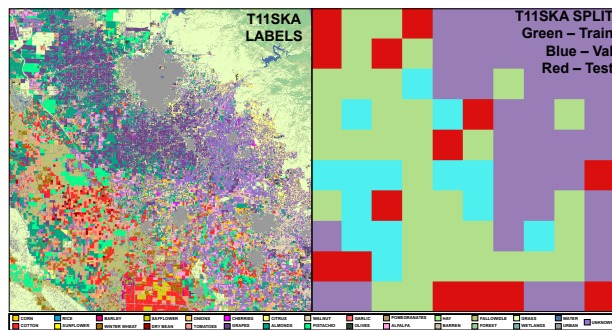
Figure 3: Visualisation of T11SKA Sentinel Tile CDL labels and grid split used in our experiments. Regions in Purple represent Unknown and were not used in analysis

4.2 Satellite Imagery We use multi-spectral satellite imagery from Sentinel-2 satellites [8] for our experiments. Specifically, we downloaded data for year 2018 from the COPERNICUS/S2_SR [2] collection on Google Earth Engine. The Sentinel-2 data product has 13 spectral bands at different spatial resolutions of 10, 20, and 60 meters. We leave out the atmospheric bands (Band 1, 9, and 10) of 60 meters resolution (due to coarse resolution) and re-sample all the remaining bands to 10 meters using the nearest neighbor method. Due to irregular temporal sampling of Sentinel-2 imagery across years we create a multi-spectral mosaic at a 15-day interval while considering cloud filters at every timestamp. This resulted in a final dataset of shape 24 x 10 x 10980 x 10980 for a year, where dimensions represents number of composite images, channels, rows, and columns respectively with each row and columns pixel denoting a 10m spatial length (Thus each satellite image's height and width is 109800m). Each band is normalised with its respective maximum and minimum. One can obtain similar mosaic sets for any year, such as 2019 or 2020.

4.3 Weather Data We used Daymet dataset as our source of weather data. Specifically, we downloaded daily weather data from NASA/ORNL/DAYMET_V4 [3] collection on Google Earth Engine. The Daymet product has seven bands namely, duration of the daylight period, daily total precipitation, incident shortwave radiation flux density, snow water equivalent, maximum 2-meter air temperature, minimum 2-meter air temperature, and average partial pressure of water vapor. Daymet is available daily at 1km spatial resolution. However, we found that downloading the data at this resolution led to a lot of missing data, so we download all bands of data daily at a resolution of 10 km

for our study region. We then resample weather data (using nearest neighbor interpolation method) to match the spatial extents with our multi-spectral satellite data and normalise each band with its respective minimums and maximums. This resulted a final shape of 365 x 7 x 10 x 10 for a year, where dimensions represents number of timestamps, channels, rows, and columns respectively with each row and columns pixel denoting a 10km spatial length.

4.4 Labels We get our labels from the Cropland Data Layer (CDL) [4] provided by the United States Department of Agriculture (USDA). USDA annually releases the CDL [4] which is a publicly available land-cover classification map for the entire country at a 30-meter resolution that includes major and minor crop commodities. Since CDL is available at 30 m resolution, we resample it to 10m using nearest neighbor to match our satellite imagery dataset. There are over 200 class types in CDL, with some irrelevant to our crop mapping task, so we follow a class combination scheme similar to that performed in CALCROP21[11] by combining the pixel labels to one of 33 classes (Please refer to Supplementary material for full list). In addition, We also have an Unknown class to denote those pixels that do not belong to our 33 classes of interest. Therefore, for our study region, the final labels are of shape 34 x 10980 x 10980 for each year.

4.5 Implementation details Like CALCROP21 [11], we adopt a grid-based splitting of the study region to ensure spatial auto-correlation does not negatively impact our evaluation. Specifically, we divide our study region into 100 grids of size 10kmx10km each (equivalent to 1098 x 1098 pixels). Furthermore, we eliminate grids that do not contain more than 50% pixels belonging to crop classes. It is also known that CDL has issues with noise and incorrect labels at the boundaries of crop fields. To address this, we perform one level of morphological erosion, i.e remove one layer of pixels at object boundaries to reduce the effect of label uncertainty at farm boundaries. We further remove small connected components (contiguous set of pixels belonging to the same class) to reduce the effect of noisy labels in the training process, please refer to [11] for diagrammatic representation of this process. Since we follow a grid-wise scheme for the satellite data, we also split the weather data into this grid scheme. This results in one weather reading per day for each grid, which is a reasonable approximation, as it can be expected for a 10km x 10km region to have very similar weather due to large spatial auto-correlation in weather patterns..

---

[2]https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR#bands

[3]https://developers.google.com/earth-engine/datasets/catalog/NASA_ORNL_DAYMET_V4

---

[4]https://nassgeodata.gmu.edu/CropScape/

5 Results

The main objective of this paper is to show the benefit of bringing in weather data for crop mapping compared to methods that use only satellite data. As mentioned, for the satellite-only based approach, we chose STATT [10] to be compared against our weather-based model.

Training Set: Following the removal of grids due to lack of crop cover, we were left with 62 grids for the T11SKA region in 2018. From these 62 grids, We randomly selected 34 grids for training, 14 for validation, and remaining 14 for testing. We keep the training and validation sets across the experiments the same. As mentioned before, each grid is of size 1098 x 1098 pixels, thus signifying a large spatial region and also each grid is consistent across time. The distribution of these grids can be seen in Figure 3. By using the grid based splitting of our study region, we avoid spatial auto-correlation between pixels (nearby pixels tend to be of same type) to negatively affect our evaluation. Furthermore, we exclude 30 pixels on each side of each grid to further remove the effect of spatial auto-correlation in our analysis. Both WSTATT and STATT were trained using these 34 grids, and the best model was chosen using the 14 validation grids. Both models were trained using learning rate of 0.0001, 50 epochs and using cross-entropy loss.

5.1 Year End Prediction In this section, we discuss the predictive performance of experiments where entire year's data is available for making the crop type prediction, i.e predicting the crop map at the end of the year with all 12 months of data available. First, we evaluate STATT and WSTATT on 14 test grids (denoted by the red boxes in Fig. 3) using data from the same year as the training data (2018). Specifically, we compare their F1 scores for different crop classes while ignoring classes with less than 100,000 pixels from this evaluation. We do not compare accuracy as it does not give a complete picture of correctness of each method, due to imbalance in classes for which F1 score is better tuned. WSTATT performed better than STATT with an overall improvement of 0.02 in average F1 score, with major improvements in Corn (0.08) and Walnuts (0.04). We further evaluate these algorithms in a more difficult setting where data for 14 test grids come from different years (2019 and 2020). The idea is to assess models' ability to generalize across time. For this setting, we observe a 0.15 increase in average F1 score for year 2019 using WSTATT compared to STATT, and an increase of 0.12 for year 2020 (Please refer to Table 1) Major improvements were seen in Almonds (0.31 increase in 2019 and 0.18 increase in 2020) and also in Alfalfa (0.25 increase in both years). Please refer to full length paper

Table 1: Comparison of WSTATT and STATT in terms of Average F1 Score for Year end predictions across various years. The numbers in bold correspond to the best across each experiment.

Year End Prediction Average F1 score		
Test Set	STATT	WSTATT
T11SKA 2018	0.8171	**0.8398**
T11SKA 2019	0.5857	**0.7365**
T11SKA 2020	0.6201	**0.7368**

for class-wise comparison between the two algorithms. As the results suggest, WSTATT is able to learn robust discriminative features among crop classes by leveraging the relationship between weather and crop-growth.

5.2 Early Prediction In this section we evaluate the performance of the models for early prediction, i.e obtaining the crop map without entire year's data. Both STATT and WSTATT have attention modules allowing dynamic timestamp weight allocation followed by aggregation. Hence, even with partial year's data, the attention module can assign weights to available timesteps and provide prediction probabilities allowing for early prediction at various time scales (6 months, 8 months, etc).

For this experiment, we vary the amount of data available for prediction as 6, 8, 10, and 12 months, where the year starts from January. Table 2 shows the early prediction results for 2019 (results for year 2020 can be found in full length paper). As we can see, WSTATT performs significantly better than STATT in the early prediction, for both years 2019 and 2020 which were not considered during training.

With just eight months of data in the test regions over both years, WSTATT consistently outperforms STATT with major differences in Cotton, Almonds, Pistachio and Grapes. This trend extends to 10 month predictions as well. Another important factor to note is that, though performance improves for both models with more data, WSTATT reaches its best score values much faster. For example, from the Cotton scores in 2019, we observe that with just eight months of data, WSTATT can achieve good accuracy, and with four more months of data, the score increases by 0.05. However, in the same timeframe, STATT's eight-month score is poor and shows a good value only when the full year's data is provided. This indicates that WSTATT is using the weather data to make earlier predictions effectively using the partial year's data.

To provide visual analysis, we show prediction maps from 2019 for certain regions. Figure 4 illustrates two such cases where the prediction using only eight months of data is close to the ground truth of CDL (shown in the last column). Although some corrections are made

Table 2: Comparison of WSTATT and STATT in terms F1 Score over various early prediction settings on test grids. The numbers in bold correspond to the best across each experiment.

| Data Provided | T11SKA 2019 Early Prediction | | | | | | | |
| | 6 MONTHS | | 8 MONTHS | | 10 MONTHS | | 12 MONTHS | |
Crop Class	STATT	WSTATT	STATT	WSTATT	STATT	WSTATT	STATT	WSTATT
Corn	**0.2747**	0.0819	0.3606	**0.6883**	0.5043	**0.7091**	0.5761	**0.7002**
Cotton	0.0064	**0.6156**	0.1934	**0.8483**	0.7590	**0.8943**	0.8888	**0.8951**
Winter_Wheat	0.0500	**0.3850**	0.2000	**0.2785**	0.2112	**0.3042**	0.2166	**0.3516**
Tomatoes	0.1618	**0.3469**	0.5890	**0.7763**	0.7485	**0.8204**	0.7790	**0.7986**
Grapes	0.0134	**0.2306**	0.0727	**0.5806**	0.4015	**0.6654**	0.6719	**0.8038**
Citrus	0.5473	**0.8148**	0.6593	**0.8041**	0.7348	**0.8285**	0.7814	**0.8471**
Almonds	0.0757	**0.2293**	0.1794	**0.7100**	0.3078	**0.7503**	0.4615	**0.7728**
Walnut	**0.2788**	0.2308	0.3341	**0.6913**	0.4889	**0.7619**	0.5187	**0.7252**
Pistachio	0.0000	**0.1142**	0.0006	**0.4340**	0.0198	**0.4682**	0.4649	**0.7261**
Alfalfa	0.2676	**0.5762**	0.3393	**0.6364**	0.4300	**0.7310**	0.4985	**0.7447**
Average	0.1676	**0.3625**	0.2928	**0.6448**	0.4606	**0.6933**	0.5857	**0.7365**
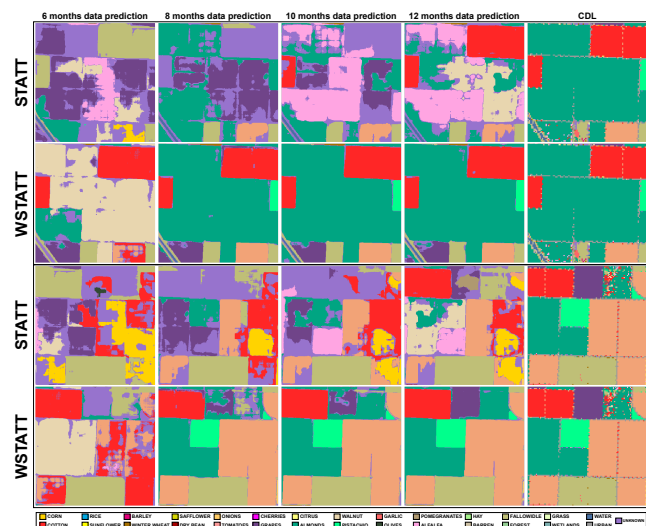
Figure 4: Cases where weather based model (WSTATT) gets a good prediction well before the end of the year

with more data, the majority of the patch is accurate. However, with the same amount of data, STATT cannot provide a good prediction and even with more data provided is unable to reach an accurate prediction. Please refer to the full length paper for a similar visual study of another region.

Through our experiment, we have found that the weather-based model can adjust to different temporal scales and produce accurate maps well before the end of the year, reaching even higher accuracy by the end of the year. This is a significant advancement in crop mapping since the CDL map for a given year is usually provided a month after the year ends, which would be around late January or early February of the next year. Our method can provide the map as early as August of the same year, which can benefit downstream tasks that rely on robust crop maps such as yield prediction.

5.3 Attention Analysis The previous sections show that the weather-based model outperforms the satellite-

only approach which can be attributed to the integration of the inverse modeling scheme in WSTATT. While both approaches utilize a feed forward attention module that assigns a weight to each timestep, STATT relies only on satellite data but WSTATT uses weather data as well. In this section, we examine the impact of including weather data on the attention module.

Some key factors about a crop are its planting date, growth pattern, biomass at maturity, harvest date, crop residue, and leaf shedding. They all can play a role in discriminating crop classes but not all of them are consistent across years. For example, planting and harvesting dates can change over years. Algorithms that identify features that are more generalizable such as the growth pattern, timing of crop maturity, etc. are more likely to perform better across different years. Since weather is tightly coupled with crop growth and maturity, it enables WSTATT to focus more on generalizable features instead of over-fitting on noisy features such as crop residue after harvesting, planting date etc.

Figure 5 shows attention weights assigned by both algorithms to each timestep for four crop classes. Each subplot displays a different crop, indicated in the top left corner, with the x-axis representing the timeframe and the y-axis showing the weightage for the corresponding timestamp. To ensure attention weights are representative of that class, we selected a patch that belong completely to that class. As we can see, attention weights are very different from both algorithms. WSTATT weights are much concentrated on few timesteps compared to STATT which suggest better understanding of discriminative periods for different classes. For instance, WSTATT focused on June to September for almonds, which was the easiest period to distinguish them from other crops. In contrast, STATT attention scores were almost equal from April to September. Similarly, for Corn, STATT focused on the period from April to July,
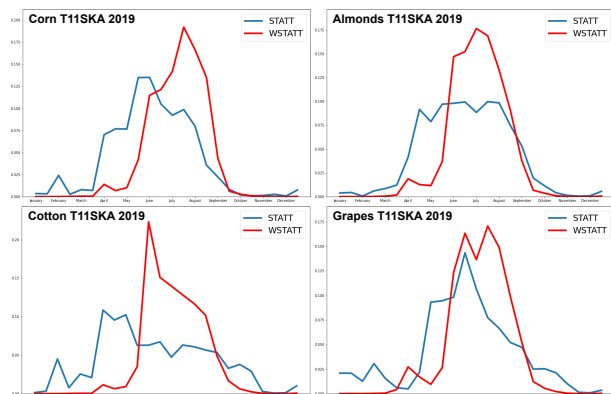
Figure 5: A diagrammatic representation of the attention weights for each crop class in 2019

which could cause confusion with winter wheat, while WSTATT paid attention to July to September.

6 Discussion on Impact of Weather

In this section, we provide domain insights into the trends observed in F1 scores for the two algorithms as reported in Table 2. Since Citrus and Alfalfa maintain their leaves throughout the year, both algorithms are providing good predictive performance even from 6 months onwards with WSTATT significantly better for these two crops even at 6 months. Grapes, a perennial crop, are harvested in September or October and shed their leaves during winter. WSTATT is quicker than STATT in detecting this, with a notable increase in performance of 8 month prediction as opposed to 10 month prediction for STATT. Almonds, pistachios, and walnuts are also perennial crops, but they shed their leaves later in the year, around October to December, causing a jump in their performance around the eight months timestamp. Pistachio, being the last to shed its leaves, shows a boost in performance only on the last timestamp.

Tomatoes have different planting times but are harvested between July and October. Therefore, their predictive performance shows a significant increase at eight months for WSTATT. Similarly, Cotton is typically planted in April, but it reaches good vegetative cover later in the year, resulting in a sharp increase in predictive performance for WSTATT but not for STATT. Corn is usually planted in May, but its leaf area index is still low after six months, making it difficult to classify accurately. At this time, winter wheat is almost at the end of its growth cycle, and some fields may have crop residue, which can be confused with corn. However, since winter wheat's growth cycle is complete, Table 2 shows that its predictive performance does not improve significantly beyond the six-month prediction, regardless of the data provided for both methods over the years.

The results indicate that including weather data is useful for identifying harvest periods more quickly and accurately capturing intervals of leaf loss. This can be attributed to the fact that inclusion of weather data allows for the model to quickly discern the difference between crop classes. A key point to note is that while higher frequency satellite imagery might offer improvements in early prediction, but the key factor determining early prediction capability is the timing of the discriminative period for a crop, irrespective of the temporal resolution.

7 Conclusion and Future Work

Our work introduces a new method for crop mapping that considers crop growth as a system influenced by physical drivers such as weather and soil type. We developed an inverse modeling approach using satellite imagery and weather data to provide accurate pixel-wise crop maps for a given area of interest. Our weather-based method outperforms traditional satellite-only methods for year end and early prediction tasks. Our early prediction allows for robust crop maps to be available to farmers and the public much sooner than the current standard (upto 5 months in advance). We found that including weather embeddings improved attention scores and allowed the model to focus on the most discriminative period.

For future work, we plan to include a generative forward modeling process in combination with the current inverse framework. Currently, our focus was on evaluating the performance over different time periods. We plan to extend the study region, and assess how weather data can enable generalization across space by comparing against existing domain adaptation techniques. Another future work direction is to include more physical parameters in the deep learning approach, such as soil information and management practices.

In conclusion, this work showed the immense benefit of bringing weather data into pixel-wise crop mapping and we believe that this work will convince remote sensing practitioners to use such an approach in their tasks, replacing the traditional methods of relying solely on satellite imagery for land cover mapping tasks.

8 Acknowledgments

References

[1] Nicoletta Addimando et al. A deep learning approach for crop type mapping based on combined time series of satellite and weather data. *IPRSIS Archives*, 2022.

452

[2] Mouad Alami Machichi et al. Crop mapping using supervised machine learning and deep learning: a systematic literature review. *IJRS*, 2023.

[3] Lukas Blickensdörfer et al. Mapping of crop types and crop sequences with combined time series of sentinel-1, sentinel-2 and landsat 8 data for germany. *RES*, 2022.

[4] Claire Boryan et al. Monitoring us agriculture: the us department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto International*, 2011.

[5] XL Cai et al. Integrating remote sensing,census and weather data for an assessment of rice yield,water consumption and water productivity in the indo-gangetic river basin. *Agricultural Water Management*, 2010.

[6] Emmanuel Capliez et al. Temporal-domain adaptation for satellite image time-series land-cover mapping with adversarial learning and spatially aware self-training. *IEEE JSTARS*, 2023.

[7] Ritesh Dash et al. Classification of crop based on macronutrients and weather data using machine learning techniques. *Results in Engineering*, 2021.

[8] Matthias Drusch et al. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *RSE*, 2012.

[9] Saskia Foerster et al. Crop type mapping using spectral–temporal profiles and phenological information. *Computers and Electronics in Agriculture*, 2012.

[10] Rahul Ghosh et al. Attention-augmented spatio-temporal segmentation for land cover mapping. In *International Conference on Big Data*. IEEE, 2021.

[11] Rahul Ghosh et al. Calcrop21: A georeferenced multi-spectral dataset of satellite imagery and crop labels. In *International Conference on Big Data*. IEEE, 2021.

[12] Rahul Ghosh et al. Robust inverse framework using knowledge-guided self-supervised learning: An application to hydrology. In *28th ACM KDD*, 2022.

[13] Patrick Helber et al. Eurosat:a novel dataset and deep learning benchmark for land use and land cover classification. *IEEE J-STARS*, 2019.

[14] Sepp Hochreiter et al. Long short-term memory. *Neural computation*, 1997.

[15] Yunfeng Hu et al. A deep convolution neural network method for land cover mapping:a case study of qin-huangdao,china. *Remote Sensing*, 2018.

[16] Sonal Jain et al. Machine learning convergence for weather based crop selection. In *IEEE SCEECS*, 2020.

[17] Shunping Ji et al. 3d convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing*, 2018.

[18] Xiaowei Jia et al. Spatial context-aware networks for mining temporal discriminative period in land cover detection. In *SDM*. SIAM, 2019.

[19] Yves Julien et al. The yearly land cover dynamics (ylcd) method: An analysis of global vegetation from ndvi and lst parameters. *RES*, 2009.

[20] Nataliia Kussul et al. Deep learning classification of land cover and crop types using remote sensing data. *IEEE GRSL*, 2017.

[21] Rose M Rustowicz et al. Semantic segmentation of crop type in africa:a novel dataset and analysis of deep learning methods. In *IEEE CVPR Workshops*, 2019.

[22] Masoud Mahdianpari et al. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. *RS*, 2018.

[23] Cristina Milesi et al. Crop yield estimation at gram panchayat scale by integrating field, weather and satellite data with crop simulation models. *Journal of the Indian Society of Remote Sensing*, 2022.

[24] Pushpa Mohan et al. Deep learning based weighted som to forecast weather and crop prediction for agriculture application. *Int. J. Intell. Eng. Syst*, 2018.

[25] Brice Mora et al. Global land cover mapping: Current status and future trends. *Land Use and Land Cover Mapping in Europe: Practices & Trends*, 2014.

[26] Luciana Nieto et al. An integrated approach of field, weather,and satellite data for monitoring maize phenology. *Scientific Reports*, 2021.

[27] Ian Olthof et al. Mapping northern land cover fractions using landsat etm+. *RSE*, 2007.

[28] Marc Rußwurm et al. End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS JPRS*, 2023.

[29] Mike Schuster et al. Bidirectional recurrent neural networks. *IEEE TSP*, 1997.

[30] Rai Schwalbert et al. Mid-season county-level corn yield forecast for us corn belt integrating satellite imagery and weather variables. *Crop Science*, 2020.

[31] Christopher Storie et al. Deep learning neural networks for land use land cover mapping. *IGARSS*, 2018.

[32] Iwan Supit et al. Assessing climate change effects on european crop yields using the crop growth monitoring system and a weather generator. *Agricultural and Forest Meteorology*, 2012.

[33] Swapan Talukdar et al. Land-use land-cover classification by machine learning classifiers for satellite observations—a review. *Remote Sensing*, 2020.

[34] Nima Teimouri et al. A novel spatio-temporal fcn-lstm network for recognizing various crop types using multi-temporal radar images. *Remote Sensing*, 2019.

[35] Gabriel Tseng et al. Cropharvest: A global dataset for crop-type classification. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

[36] Junshi Xia et al. Openearthmap: A benchmark dataset for global high-resolution land cover mapping. In *Proceedings of IEEE WACV*, 2023.

[37] SHI Xingjian et al. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *NeurIPS*, 2015.

[38] Le Yu et al. A multi-resolution global land cover dataset through multisource data aggregation. *Science China Earth Sciences*, 2014.

[39] Hongwei Zhao et al. Evaluation of five deep learning models for crop type mapping using sentinel-2 time series images with missing information. *RS*, 2021.