A geospatially conditioned feature optimization approach for remote sensing-based maize yield and water productivity estimations across Kansas and Colorado

by

Ross Wesley Unruh

B.S., Kansas State University, 2021

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Approved by:

Co-Major Professor
Dr. Romulo Lollato

Approved by:

Co-Major Professor
Dr. Raj Khosla

Copyright

Abstract

Accurate and efficient assessments of crop production and agricultural water use are critical to the sustainability of national and global food systems, economic stability, and water resource management. Remote sensing technologies, particularly radar and optical remote sensing data from spaceborne satellite systems, have been successfully utilized for gathering data required to make these assessments. However, there exists a knowledge gap regarding the synergistic use of the data derived from these methods to account for the climatic and geospatial variations in agricultural regions.

The objectives of this study were to 1) determine the optimal temporal acquisition windows, focal statistic neighborhood dimensions, and vegetation indices, for optical and radar remote sensing data's use in estimating maize yield across climate zones in Kansas and Colorado, and 2) identify environmental and climatic characteristics of high and low water productivity areas using OpenET data integrated with geospatially optimized yield models and weather station data.

Nine site-years resulting from five maize fields and four years spanning across the states of Kansas and Colorado in the U.S. were analyzed. Site-years represented varying climatic conditions and geospatial properties. Seasonal imagery from Sentinel-1 radar and Sentinel-2 optical satellite systems was acquired via Google Earth Engine. ArcGIS Pro was used to generate feature sets for five focal statistic neighborhood dimensions (1x1, 3x3, 5x5, 7x7, and 9x9) and ten vegetation indices and bands (RVI, DpRVI, VH/VV, VH, VV, NDVI, NDRE, SAVI, GNDVI, and EVI). The resulting features for each site were then statistically assessed and selected based on their correlation with the yield data, and subsequently incorporated into a random forest regression model. The final models were evaluated for their ability to accurately

estimate yield, both intra- and inter-site-year. Results showed that the feature selection process favored diversity over uniformity in vegetation index and focal statistic neighborhood dimension optimizations. Additionally, intra-site-year random forest regression-based yield predictions performed better than the inter-site-year predictions for all but one site-year, highlighting the positive impact of optimized features for specific geospatial and climatic regions.

These models were then implemented at the county scale and compared with published USDA-NASS county-average maize yield predictions. Additionally, the USDA-NASS reported county average yield estimates were used in conjunction with OpenET, a remote-sensing-based evapotranspiration dataset, to estimate county-scale water productivity for a total of 32 site-years across 17 counties in Kansas and Colorado. Statistical analysis of each site-year's water productivity was conducted to determine its correlation with several climatic variables, including air temperature, relative humidity, precipitation, wind speed, soil temperature, solar radiation, elevation, and SSURGO soil data. The findings from this study indicated that the models developed in Chapter 1 generally did not perform well when scaled to the county level but were able to generate multiple county-scale yield predictions within 1 standard deviation of the USDA-NASS reported county average yield estimates. Low average daily wind speed was identified as the most impactful environmental metric for characterizing areas with high water productivity.

Table of Contents

List of Figures

List of Tables

Acknowledgements

The successful completion of this thesis is a direct result of the dedication and support I received from each past and present member of the Khosla Precision Agriculture Lab at Kansas State University. I am sincerely grateful for the guidance, assistance, and friendship of Drs. Jeff Siegfried, Dipankar Mandal, Wubengeda Yilma, Atin Majumder, and the soon-to-be Dr. Raghav Joshi. I am especially grateful for the mentoring and professional coaching I received from Dr. Raj Khosla, who provided me with the opportunity to achieve a master's degree.

I am honored to have worked alongside and been helped by the administrative, faculty, and research staff of the Kansas State University Agronomy Department and the Colorado State University Department of Soil and Crop Sciences, whose efforts made this research possible. Additionally, I am thankful for the important research being conducted by the Artificial Intelligence for Sustainable Agriculture project, and the people that make it possible.

Finally, I am thankful to my research committee members, Drs. Dipankar Mandal, Shawn Hutchinson, Romulo Lollato, and Raj Khosla, for their willingness to work across multiple time zones and countries to help me accomplish my degree program.

Dedication

This thesis is dedicated to my mother, Dr. Jana Unruh, who encouraged and inspired me to pursue graduate school, and to my father, Jim Unruh, who relentlessly supported me throughout the entire process. I also dedicate this thesis to my siblings, Grant, Jill, and Faith, whose love and support helped me persevere. Finally, this thesis is dedicated to my late grandparents Betty and Eldon Huffman, and Arch and Mary Unruh, who instilled in me a passion for curiosity.

Chapter 1 - Estimating Field-Scale Maize Yield with Geospatially Conditioned Remote Sensing Feature Optimizations

1.1 Introduction

1.1.1 Global Importance of Maize

The Food and Agriculture Organization (FAO) of the United Nations reported that from 2010 to 2024, more maize (*Zea mays* L.) was produced globally than any other cereal crop (Food and Agriculture Organization of the United Nations, 2025). In 2024 alone, 1.22 billion tons of maize were harvested around the world from 254.5 million hectares of farmland, an area that, if combined, for reference, would be slightly larger than the country of Algeria, the tenth largest country in the world (Food and Agriculture Organization of the United Nations, 2025; United Nations Development Programme (UNDP), 2026). Primarily grown for human and livestock consumption, maize is a versatile crop that can be consumed with minimal processing, ground into flour, pressed for oil, and fermented for alcohol and ethanol, among many other uses. Due to its adaptability to a wide range of climates and its vast range of food and industrial uses, maize plays a crucial role in the global agricultural economic system. The Foreign Agriculture Service (FAS) of the United States Department of Agriculture (USDA) reported that globally, between 30 and 54 billion dollars were generated from maize production annually across 2020 and 2024 (Foreign Agriculture Service, 2024). For instance, monthly global maize yield estimates generated by the FAS and the National Agriculture Statistic Service (NASS) directly inform the monthly World Agriculture Supply and Demand Estimates (WASDE) reports. These reports subsequently influence entire industries existing to generate the fuel and resources needed to power thousands of dry bulk carriers, freight trains, and shipping trucks that are operated by grain trade networks (Agricultural Marketing Service, 2024). Maize production also drives governments to invest billions of dollars in research and development grants, resulting in scientific discoveries that increase yield and sustainability practices across the world (Plastina & Townsend, 2023). Evidently, it is imperative that production assessments and yield predictions for this vital commodity are provided accurately and timely to economists, policymakers, growers, buyers, exporters, and researchers alike.

1.1.2 Optical and Synthetic Aperture Radar Remote Sensing of Maize

The challenge of generating accurate global maize yield predictions in a timely and accurate manner has seen significant progress with the use of space-borne remote sensing methods (Zhao et al., 2024). One such method includes the use of optical sensors to measure the reflectance of discrete spectral wavelengths from crop canopies, allowing for the monitoring of several biophysical parameters in crops (Zhao et al., 2024). Alternatively, research into the use of Synthetic Aperture Radar (SAR) remote sensing for collecting data on the physical structure of developing crops, as well as soil and canopy moisture, has resulted in the collection of useful insights into crop growth and development (Steele-Dunne et al., 2017). Unlike optical remote sensing, SAR utilizes larger, centimeter-scale wavelengths capable of penetrating cloud cover while imaging the Earth's surface (Gupta et al., 2022). Therefore, the synergistic use of these two methods of agricultural remote sensing has been identified as a promising approach to collecting gapless time-series datasets for use in agricultural yield prediction (Ghosh et al., 2022).

While many useful space-borne systems exist, the European Space Agency’s Sentinel-1 SAR and Sentinel-2 optical satellite systems are especially favorable for collecting this data due to their identical 10 m pixel spacing, relatively high temporal resolutions (6-12 days), and open source availability (European Space Agency Earth Observation Portal, 2026a, 2026b). Despite valuable developments in the use of these two remote sensing datasets, there exists a knowledge gap regarding the synergistic use of SAR and optical remote sensing to account for the climatic and geospatial variations across agricultural regions.

1.1.3 Accounting for Geospatial and Climatic Diversity in Remote Sensing-Based Maize Yield Prediction

Due to its efficient utilization of water, adaptability to a wide range of environments, and ability to grow successfully in most soil types, maize is produced in a myriad of geospatially and climatically diverse regions around the world (Food and Agriculture Organization of the United Nations, 2026). For instance, the USDA's National Agriculture Statistics Service (NASS) reported in 2024 that 36.79 million hectares of maize was harvested across each of the forty-eight contiguous states within the United States (US), spanning nine unique climatically consistent regions (National Agricultural Statistics Service, n.d.; National Centers for Environmental Information, n.d.). Major diversities in temperature, precipitation, humidity, elevation, and many more unique climatic characteristics are present across these regions, yet maize continues to be the US's second most exported agricultural commodity (Foreign Agriculture Service, 2024). Consequently, there exists significant heterogeneity in planting dates, farming practices, and genetic diversity in maize production operations. Therefore, optimizing remote sensing-based yield predictions by accounting for climatic and geospatial diversity might provide a deeper insight into the synergistic use of SAR and optical remote sensing for maize yield predictions. Three such remote sensing optimizations include vegetation indices, a spatial analysis technique called focal statistics, and the identification of the most important temporal windows for acquiring remote sensing datasets.

Since the inception of optical satellite remote sensing, researchers have been developing vegetation indices that integrate measured reflectance across multiple discrete spectral wavelengths to assess specific crop attributes (Rouse et al., 1974; Zhao et al., 2024). Likewise, SAR based vegetation indices have been developed to collect useful agronomic information by incorporating the measurements of reflectance from vertically and horizontally polarized microwave pulses against soil and crop canopies (Brisco et al., 1992; Steele-Dunne et al., 2017).

The focal statistics remote sensing optimization involves the statistical smoothing of pixel values based on a predefined focal statistic neighborhood dimension (Esri, 2025b; Luo et al., 2020). This technique, investigated by Luo et al. (2020), mitigates anomalous pixel values in Sentinel-1 and Sentinel-2 datasets by modifying the value of each pixel to take on the average value of the surrounding pixels within a predefined neighborhood window (Luo et al., 2020). Due to Sentinel-1 SAR's sensitivity to water and metallic objects such as tractors and irrigation systems, the presence of anomalous pixel values, which could negatively impact yield predictions, is a valid concern, making focal statistics a promising optimization. Additionally, incorporating different combinations of vegetation indices and focal statistic neighborhood dimensions may improve yield prediction accuracy.

Finally, given the large number of satellite remote sensing datasets acquired within a given maize growing season, identifying the temporal windows that contain the most impactful datasets could enable a more streamlined, less resource-intensive approach to generating accurate maize yield predictions. Additionally, identifying these temporal acquisition windows could lead to useful discoveries regarding how vegetation indices and focal statistic neighborhood dimension optimizations perform within discrete phenological maize growth stages.

Therefore, the objectives of this study were to 1) determine the optimal temporal acquisition windows, focal statistic neighborhood dimensions, and vegetation indices for radar and optical remote sensing data's use in estimating maize yield across climate zones in Kansas and Colorado, and 2) determine whether the optimization of remote sensing datasets for specific regions leads to more accurate maize yield predictions.

1.2 Materials and Methodology

1.2.1 Experimental Locations and Areas of Interests

Nine site-years resulting from five maize fields and four years located across the U.S. states of Colorado (CO) and Kansas (KS) were studied. Fields represented varying climatic and geospatial properties (see Table 1.1 and Figure 1.1), since they were located across a wide longitudinal gradient of elevation, and of average temperature and accumulated precipitation in the maize production months of May through October (see Figure 1.2, 1.3).

Table 1.1 Location, identification, area, and latitude and longitude of the nine site-years used in this study.

Location	Site-Year ID	Area (ha)	Latitude	Longitude
Topeka, KS	TO_2021	3.34	39.0762869	-95.7699724
Rossville, KS	RO_2022	7.40	39.1181554	-95.9269275
Rossville, KS	RO_2023	2.68	39.1181554	-95.9269275
Colby, KS	CO_2024	2.07	39.3859843	-101.072515
Ft. Collins, CO	FC_2021	2.38	40.6652105	-104.998082
Ft. Collins, CO	FC_2024	3.39	40.6652105	-104.998082
Fruita, CO	FR_2022	3.58	39.1814185	-108.697746
Fruita, CO	FR_2023	1.90	39.1814185	-108.697746
Fruita, CO	FR_2023	2.30	39.1814185	-108.697746

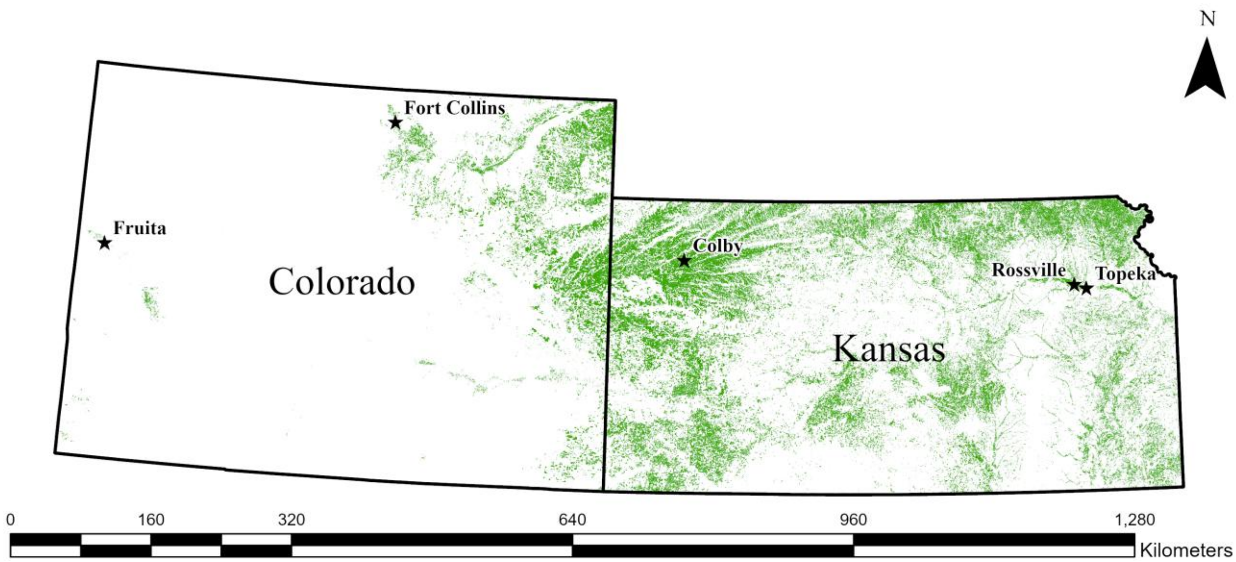


Figure 1.1 Study site locations across Colorado and Kansas, USA, with 2021-2024 maize fields displayed in green (USDA Economic Research Service, 2025).

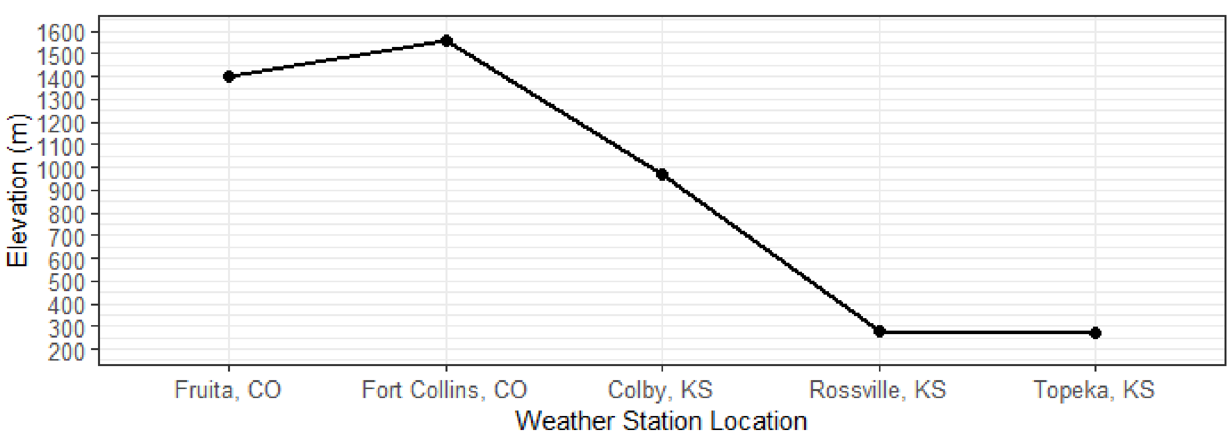


Figure 1.2 Elevation across study sites based on Kansas State University Mesonet and Colorado State University CoAgMET weather station location (Colorado Agriculture and Meteorological Network, 2026; Kansas Mesonet, 2026).

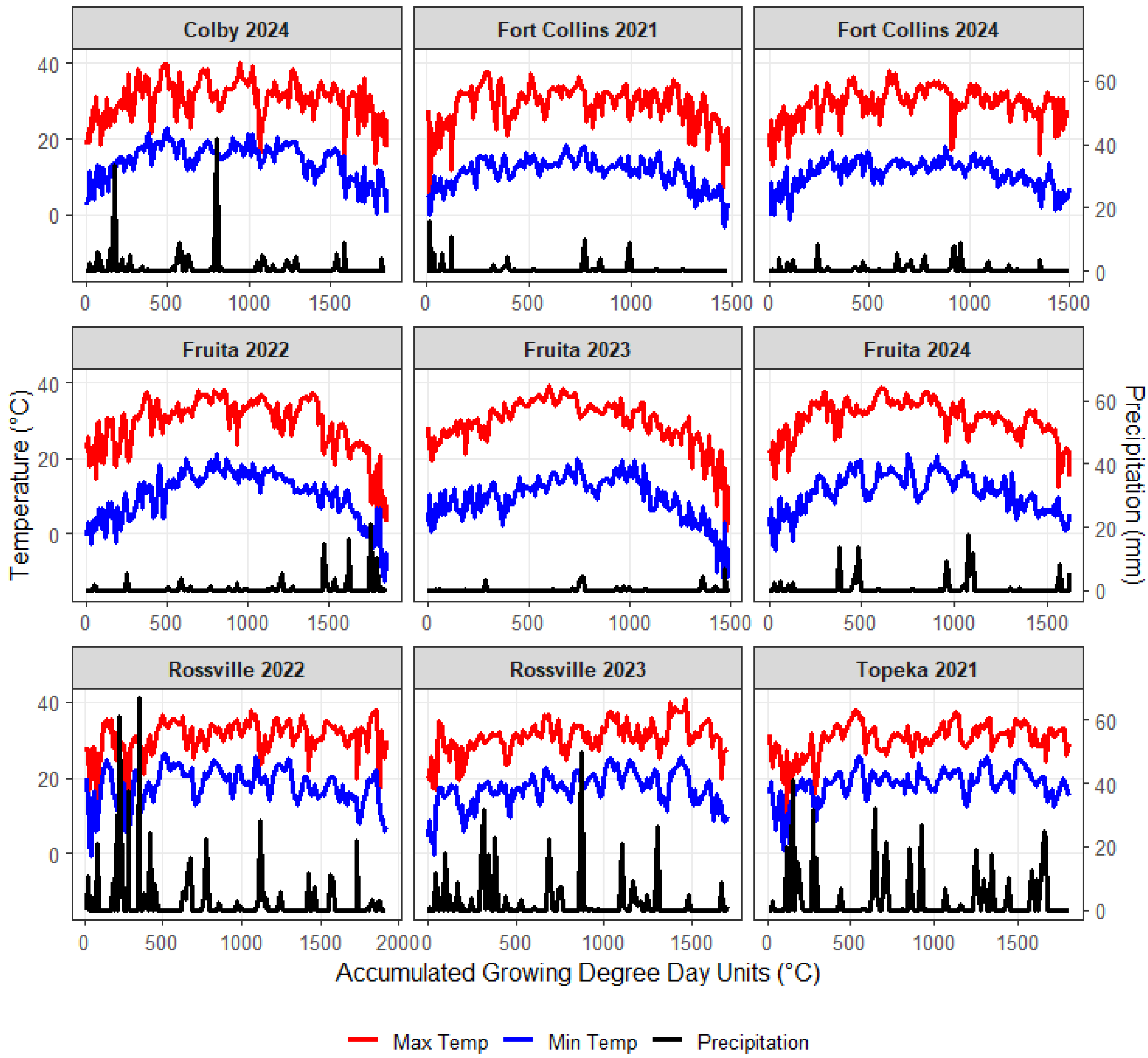


Figure 1.3 Plots displaying the minimum (blue) and maximum (red) temperatures, and precipitation (black) from planting to harvest for each site-year derived from Kansas State University Mesonet and Colorado State University CoAgMET weather stations (Colorado Agriculture and Meteorological Network, 2026; Kansas Mesonet, 2026).

Areas of Interest (AOIs) for each site-year were selected from portions of maize fields that received uniform management practices. Management decisions such as tilling, seeding rate, seed variety, irrigation, and nitrogen fertilizer application varied across site-year AOIs based on differences in geographically specific standard practices (see Table 1.2). Further, the maize fields used in this study were part of a broader, multi-year nitrogen and irrigation use efficiency study. This resulted in some AOI's receiving nitrogen applications (ranging from 68 to 90.72 kg ha^{-1}) while others were left unfertilized to mitigate the presence of residual nitrogen in the soil for the following years study. Irrigation applications were generally applied via linear and center-pivot irrigation systems, however, the Fruita 2022 site-year utilized furrow irrigation. Irrigation rates were influenced by precipitation, resulting in variations across site-years. Conventional tillage practices were used in all but the Fruita site-years, which practiced no-till. Finally, seed variety and population varied across site-years, and were determined based on site-specific environmental factors.

Table 1.2 Planting, fertilizer, irrigation, and tillage management practices across site-years.

Site-Year ID	Planting Date	Applied N (kg ha^{-1})	Irrigation Type	Seeding Rate (seeds ha^{-1})	Tillage Type
TO_2021	04/26/21	68.00	Linear	79,040	Conventional
RO_2022	04/22/22	68.00	Linear	79,040	Conventional
RO_2023	04/26/22	81.65	Linear	81,545	Conventional
CO_2024	05/09/24	0.00	Linear	69,189	Conventional
FC_2021	05/07/21	0.00	Center-pivot	84,016	Conventional
FC_2024	05/04/21	0.00	Center-pivot	84,016	Conventional
FR_2022	04/26/22	90.72	Furrow	79,074	No-Till
FR_2023	05/10/23	68.00	Linear	79,074	No-Till
FR_2023	04/29/24	0.00	Linear	79,074	No-Till

1.2.2 Determination of Satellite Imagery to Download

Meteorological data from local weather stations hosted by Kansas Mesonet and the CoAgMet were employed to determine specific growth stages throughout the season based on the planting dates at each site-year (see Table 1.3) (Kansas Mesonet, 2025, Colorado State University, 2025). A date range between estimated emergence and physiological maturity of maize was calculated for each site-year, and only the satellite images acquired within those date ranges were downloaded so that bare soil and physiologically mature canopy images were omitted from the datasets.

Table 1.3 Estimated growth stages based on accumulated GDD units from planting date.

Accumulated GDD Units (°C)	Growth Stage
69-264	Emergence
264-639	Tassel initiation
639-778	Tassel emergence
778-1069	Silking
1069-1361	Dough stage
1361-1500	Dent stage
1500+	Physiological Maturity

Hoeft et al., 2000; Kansas State University Agricultural Experiment Station And Cooperative Extension Service, 2007

1.2.3 Acquisition of Satellite Imagery

The European Space Agency's (ESA) Sentinel-1 and Sentinel-2 satellite systems were used in this study for their high temporal (6–12 day return time) and spatial (10 m x 10 m pixel spacing) resolutions. Google Earth Engine (GEE), an interactive website used for accessing and downloading satellite images, was used to download satellite imagery restricted to the defined site-year AOIs (Google Earth Engine, 2025). From GEE, preprocessed Sentinel-1 C-Band interferometric wide swath ground range detected Synthetic Aperture Radar (SAR) images containing both Vertical-transmit, Vertical-receive (VV) and Vertical-transmit, Horizontal-receive (VH) bands were downloaded. Likewise, preprocessed Sentinel-2 Level-2A orthorectified atmospherically corrected surface reflectance images were downloaded from GEE in Blue (490nm), Green (560nm), Red (665nm), Red Edge (705nm), and Near Infrared (NIR) (842nm) bands. Copernicus Browser, an ESA-developed website that allows for the exploration of Sentinel satellite imagery, was accessed to identify and exclude any Sentinel-2 images that contained greater than or equal to 30% cloud cover over each site-years' AOI (European Space Agency, 2025).

1.2.4 Determination of Maize Yield and Data Extraction

At each site-year, maize was harvested with a combine harvester outfitted with a grain yield monitor. Yield Editor 2.0.7, (Sudduth & Drummond, 2007), was used to clean and improve the accuracy of the raw yield monitor data. Next, ArcGIS Pro (Esri, 2025g), was used to generate point features containing the coordinates and dry yield values from the cleaned yield data. Predictive simple kriging interpolation, using the covariance function and the Stable model type within the ArcGIS Pro Geostatistical Wizard tool was executed with the generated point features to create yield rasters (Esri, 2025c). The ArcGIS Pro Aggregate tool was then used to alter the pixel spacing of the rasters to match the 10 m pixel spacing of the satellite images, with the mean aggregation technique (Esri, 2025a). Next, the Raster to Point tool in ArcGIS Pro was used to generate center-of-pixel point features from the aggregated yield rasters (Esri, 2025d). Finally, the generated center-of-pixel point features, in conjunction with the Sample tool in ArcGIS Pro, were utilized to extract the yield data and associated latitude and longitude coordinates from the aggregated yield rasters (Esri, 2025f). See Figure 1.4 for a simplified diagram of the maize yield data extraction process.

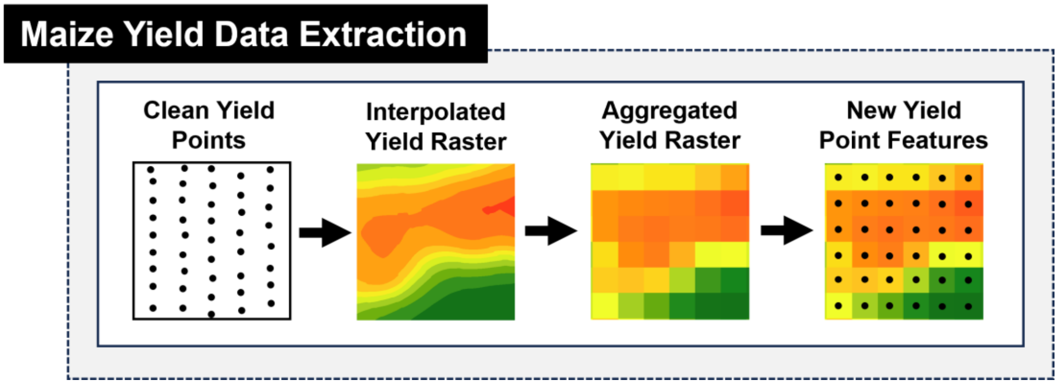


Figure 1.4 Diagram showing the process of extracting yield data from rasters generated from interpolated cleaned yield monitor data.

1.2.5 Feature Generation - Vegetation Indices

Ten vegetation indices, representing thoroughly studied and proven relationships between reflectance measurements and crop development, were chosen for the Sentinel-1 SAR and Sentinel-2 optical image sets (Table 1.4) (Barnes et al., 2000; Blaes et al., 2006; Gitelson et al., 1996; A. Huete et al., 2002; A. R. Huete, 1988; Mandal et al., 2020; Nasirzadehdizaji et al., 2019; Rouse et al., 1974; Steele-Dunne et al., 2017). For the Sentinel-1 SAR images, VH, VV, VH/VV, the Radar Vegetation Index (RVI), and the Dual-Pol Radar Vegetation Index (DpRVI) were used. For the Sentinel-2 Optical images, the Normalized Difference Vegetation Index (NDVI), Normalized Difference Red Edge (NDRE) index, Soil Adjusted Vegetation Index (SAVI), Green Normalized Difference Vegetation Index (GNDVI), and Enhanced Vegetation Index (EVI) were used.

Table 1.4 Summary of Sentinel-1 and Sentinel-2 vegetation indices/bands used to generate model input features.

Index/Band	Formula	Source
Vertical Transmit, Horizontal Receive (VH)	VH = σ_{VH^0}	Steele-Dunne et al., 2017

Vertical Transmit; Vertical Receive (VV)	VV = σ_{VV^0}	Steele-Dunne et al., 2017
VH/VV Ratio (VH/VV)	VH/VV Ratio = $\frac{\sigma_{VH^0}}{\sigma_{VV^0}}$	Blaes et al., 2006
Radar Vegetation Index (RVI)	RVI = $\frac{4\sigma_{VH^0}}{\sigma_{VV^0}+\sigma_{VH^0}}$	Nasirzadehdizaji et al., 2019
Dual-Pol Radar Vegetation Index (DpRVI)	DpRVI = $1-(m\beta)$, where $q = \frac{\sigma_{VH^0}}{\sigma_{VV^0}}, m = \frac{(1-q)}{(1+q)}$, and $\beta = \frac{1}{(1+q)}$	Mandal et al., 2020
Normalized Difference Vegetation Index (NDVI)	NDVI = $\frac{NIR-RED}{NIR+RED}$	Rouse et al., 1974
Normalized Difference Red Edge Index (NDRE)	NDRE = $\frac{NIR-RED\ EDGE}{NIR+RED\ EDGE}$	Barnes et al., 2000
Soil Adjusted Vegetation Index (SAVI)	SAVI = $\frac{NIR-RED}{NIR+RED+0.428} \times (1+0.428)$	Huete, 1988
Green Normalized Difference Vegetation Index (GNDVI)	GNDVI = $\frac{NIR-GREEN}{NIR+GREEN}$	Gitelson et al., 1996
Enhanced Vegetation Index (EVI)	EVI = $2.5 \times \frac{NIR-RED}{(NIR+6\times RED-7.5\times BLUE)+1}$	Huete et al., 2002

1.2.6 Feature Generation - Focal Statistic Windows

The Focal Statistics tool in ArcGIS pro was used to generate new features with 3x3, 5x5, 7x7, and 9x9 focal statistic neighborhood windows for every image at each site (Esri, 2025b). This tool replaces the value of each pixel with the average of the pixels surrounding it included in the selected focal statistic window dimension. This method is effective at smoothing out anomalous pixel values and reducing error (Luo et al., 2020).

1.2.7 Feature Generation - Feature Optimizations and Data Extraction

ArcPy, a Python module allowing for the automation of geospatial processing tasks in ArcGIS Pro, was used to generate each feature and extract the new data for use in the feature selection section (Esri, 2024). First, a script was run to reproject the images to match the projection of the site-year AOI they were downloaded for, standardizing the pixel locations across each image (Esri, 2025e). When reprojecting the Sentinel-1 imagery, the script converted the value of each pixel from a decibel (dB) scale to a linear scale for ease of analysis. Next, a unique script for each satellite data source was run to apply vegetation index optimizations to the image sets, yielding a new set of features for each site-year. A third script then applied the focal statistic neighborhood window dimension optimizations to each of the newly generated features (Esri, 2025b). The center-of-pixel yield data point features generated in section 1.2.4 were then used in a final script to sample the pixel values from each feature and export the data into table files to be used for analysis in R Studio, an interactive statistics software (see Figure 1.5 for a simplified diagram of the feature generation process) (Esri, 2025f).

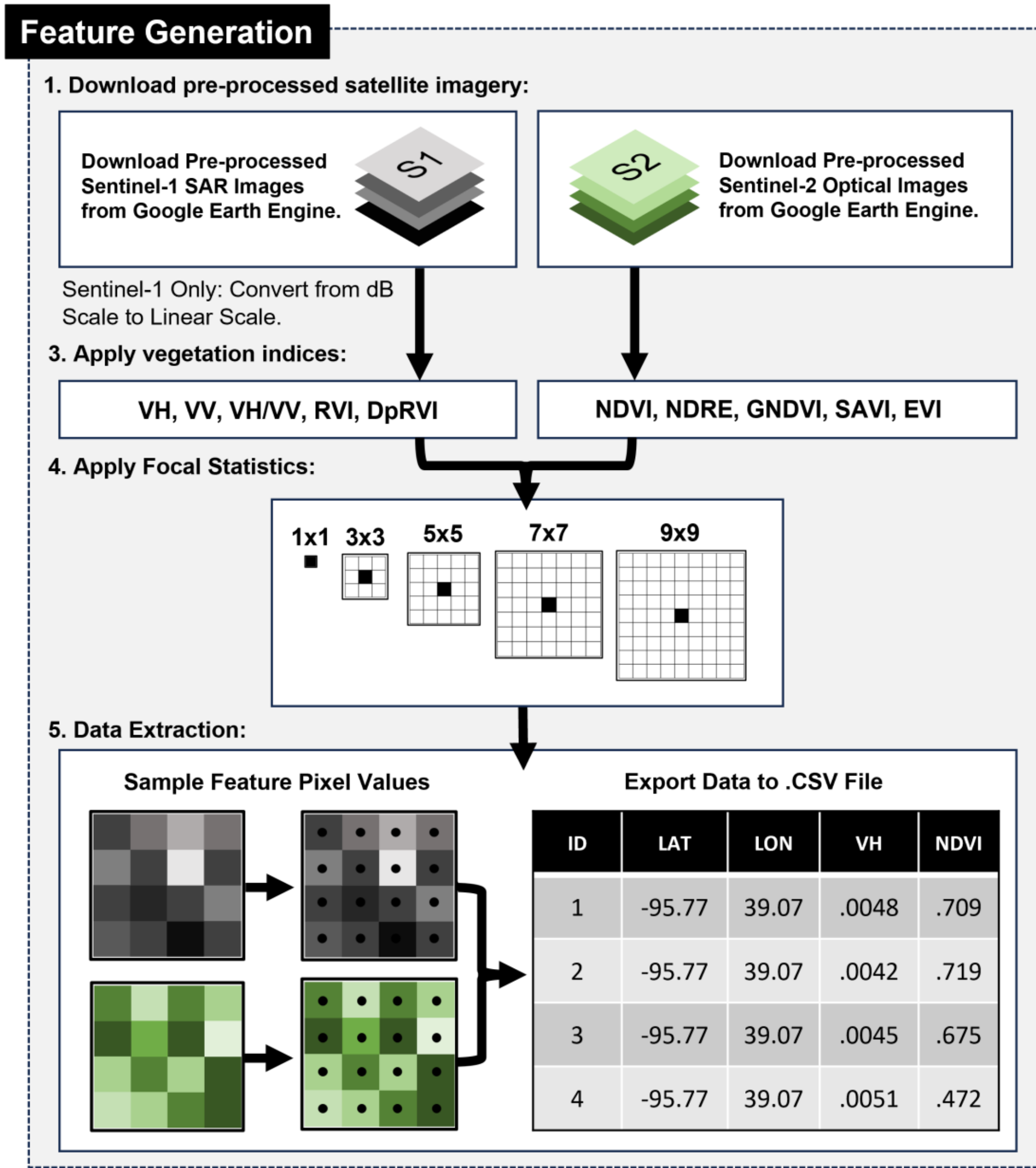


Figure 1.5 Diagram showing the process of generating and extracting data from model input features derived from Sentinel-1 and Sentinel-2 satellite imagery.

At the conclusion of the feature generation operation, a total of twenty-five (25) features for each acquired satellite image were generated: one set of original processed images with applied vegetation index optimizations, and one new set for each additional focal statistic window dimension optimization (see Table 1.5).

Table 1.5. Features generated from each Sentinel-1 and Sentinel-2 image acquisition after application of vegetation indices and focal statistic window optimizations.

		Generated Sentinel-1 Features		
VH	VH 3x3	VH 5x5	VH 7x7	VH 9x9
VV	VV 3x3	VV 5x5	VV 7x7	VV 9x9
VH/VV	VH/VV 3x3	VH/VV 5x5	VH/VV 7x7	VH/VV 9x9
RVI	RVI 3x3	RVI 5x5	RVI 7x7	RVI 9x9
DpRVI	DpRVI 3x3	DpRVI 5x5	DpRVI 7x7	DpRVI 9x9
		Generated Sentinel-2 Features		
NDVI	NDVI 3x3	NDVI 5x5	NDVI 7x7	NDVI 9x9
NDRE	NDRE 3x3	NDRE 5x5	NDRE 7x7	NDRE 9x9
SAVI	SAVI 3x3	SAVI 5x5	SAVI 7x7	SAVI 9x9
GNDVI	GNDVI 3x3	GNDVI 5x5	GNDVI 7x7	GNDVI 9x9
EVI	EVI 3x3	EVI 5x5	EVI 7x7	EVI 9x9

1.2.8 Feature Selection

Reducing the amount of input features in remote sensing-based yield estimation models via a random forest ranking approach can significantly improve model performance (Marques Ramos et al., 2020). The feature selection strategy implemented in this study modifies Marques Ramos's approach by choosing the most important feature from predefined sets of growth stages, rather than selecting the best three features from the entire feature set. This allows for a deeper understanding of both the optimizations and the importance of features at different growth stages across site-years. Additionally, this approach mitigates the potential for model overfitting and multicollinearity that could occur from using the entire feature set for each site-year. For each site-year, the Random Forest plugin for R Studio was used to generate feature importance values via the Percent Increase in Mean Square Error (%incMSE) metric for the combined Sentinel-1 and Sentinel-2 features (Liaw & Wiener, 2002; R Core Team, 2023). Using the accumulated GDD unit associated with the acquisition date of each feature, the features were organized into six (6) physiological growth stages: emergence, tassel initiation, tassel emergence, silking, dough stage, and dent stage. Within each growth stage, the feature with the highest %incMSE value was selected for use in the final model (See Figure 1.6.).

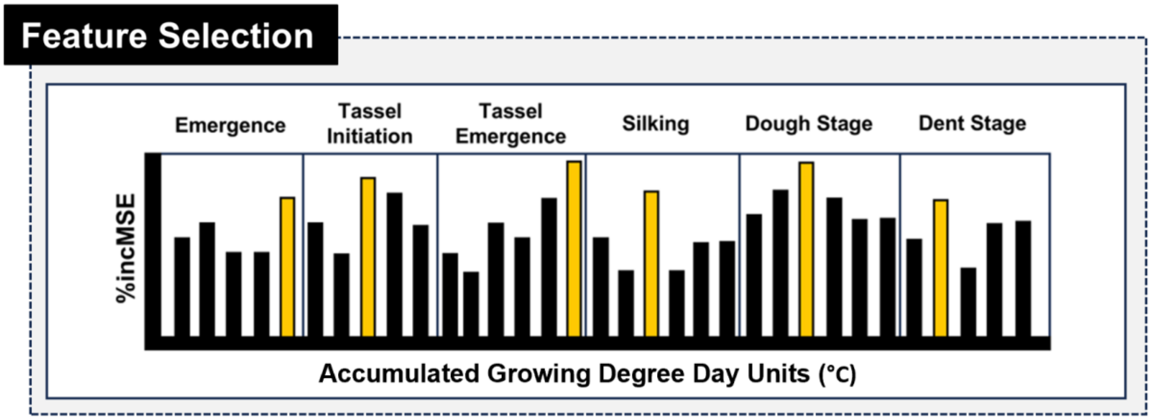


Figure 1.6 Simplified diagram demonstrating the selection of features with the highest %incMSE values from each GDD based growth stage.

1.2.9. Model Creation and Data Analysis

Intra-site-year analysis was conducted using 10-fold cross validation random forest regression in R Studio to calculate yield predictions for each site-year using the six selected features identified earlier. Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2) metrics were calculated for each site-year's yield prediction result. Using these metrics, scatterplots and visual maps displaying the observed and estimated yield were generated for use in the results and discussion section.

To determine whether the feature optimizations uniquely improved the yield predictions of the site-year regions they were selected for, inter-site-year analysis was conducted by testing each iteration of feature optimizations from one site-year with the generated features of another. This was accomplished by using the accumulated GDD unit value, vegetation index, and focal statistic window dimension optimizations for each growth stage from one site-year to select the closest available feature from the entire generated feature set of another site-year. In the event of a feature optimization not being able to be applied to the feature set of a different site-year due to differences in availability of acquisitions during a specific growth stage, that feature was omitted from the inter-site-year feature set. Each of the seventy-two (72) generated inter-site-year feature sets were then run through the same 10-fold cross validation random forest regression Model used for the intra-site-years. RMSE and R^2 metrics were calculated for each yield prediction result, and a bar chart was generated to visualize the model performance results across all site-years for each site-years optimization.

1.3 Results and Discussion

1.3.1 Feature Generation Results

Total Sentinel-1 and Sentinel-2 acquired images across site-years ranged from 25 to 127 (Table 1.6). From these images, a range of 625 to 3,175 features were generated (Table 1.6). The significant variation in acquired images and subsequently generated features can be attributed to multiple environmental and logistical factors.

Differences in the geographic locations of each site-year resulted in substantial variability in the amount of images acquired. Total days between estimated emergence and physiological maturity ranged from 112 to 150, allowing for more image acquisitions for the Fort Collins and Fruita site-years (129 to 150 days), compared to the Rossville, Topeka, and Colby site-years (112 to 119 days) (Table 1.6). This variation in days between estimated emergence and physiological maturity can be attributed to a slower accumulation of GDD units in the Colorado site-years, which had higher elevations, cooler average minimum daily air temperatures, and less precipitation than the Kansas site-years (Figures 1.2 and 1.3).

Further, some site-years had especially cloudy seasons resulting in limited Sentinel-2 acquired image useability. This was the case for the Topeka 2021, Rossville 2022 and 2023, and Fruita 2022 site-years, where 62% or less of the total acquired images were useable after filtering out images with $\geq$ 30% cloud cover (see Table 1.6).

Finally, variations in the amount of image acquisitions can be attributed to some site-year AOIs being in locations where the footprint of Sentinel-2 satellite overpasses overlapped each other, resulting in more than one image being acquired in a single day in some cases. This especially contributed to the differences in total image acquisitions over the Fort Collins site-years compared to all other site-years (see Table 1.6.).

Table 1.6 GDD based emergence and maturity dates, accumulated GDD units, total acquired Sentinel-1 and Sentinel-2 images, and generated features.

	TO_2021	RO_2022	RO_2023	CO_2024	FC_2021	FC_2024	FR_2022	FR_2023	FR_2024
Estimated Emergence Date	05/02/21	05/01/22	05/07/23	05/18/24	05/22/21	05/17/24	05/07/22	05/19/23	05/11/24
Estimated Maturity Date	08/23/21	08/22/22	08/27/23	09/14/24	10/19/21	10/09/24	09/14/22	09/30/23	09/16/24
Accumulated Days	114	113	112	119	150	146	131	135	129
Accumulated GDD Units (°C) by Estimated Maturity Date	1,487.31	1,496.14	1,491.32	1,498.81	1,472.39	1,496.78	1,499.06	1,496.39	1,493.39
Accumulated GDD Units (°C) Per Day	13.05	13.24	13.32	12.60	9.82	10.25	11.44	11.08	11.58
Acquired Sentinel-2 Images	23	23	23	24	59	54	26	27	23
Useable Sentinel-2 Images After Cloud Filter	12	11	8	20	45	40	16	22	12
Percentage of Useable Sentinel-2 Images	52%	48%	35%	83%	76%	74%	62%	81%	52%
Sentinel-2 Generated Features	600	550	400	500	2,250	2,000	400	550	600
Acquired Sentinel-1 Images	8	7	9	36	37	36	10	23	8
Sentinel-1 Generated Features	200	175	225	900	925	900	250	575	200
Total Acquired Images	32	29	25	56	127	116	26	45	32
Total Generated Features	800	725	625	1,400	3,175	2,900	650	1,125	800

1.3.2. Feature Selection Results

At the conclusion of the feature selection process, the vegetation index and focal statistic window dimension features optimizations varied widely across site-years and growth stages, however there were some notable trends (see Table 1.7 and Figure 1.7).

There were no site-years that included one single feature optimization, rather, the selection process resulted in a diverse set of vegetation index and focal statistic window dimension feature optimizations across growth stages for each site-year. Additionally, in all but the Fruita 2022 site-year, the feature selection process generated a mix of Sentinel-1 and Sentinel-2 based features across growth stages. This result is significant because it demonstrates that in all instances of this study, a diverse set of feature optimizations were preferred over one single optimization across growth stages, and further, more accurate yield predictions are generated when a diverse set of feature optimizations are used.

Across all site-years, Sentinel-1 based features were selected more frequently for the tassel initiation and tassel emergence stages, comprising 77% and 55% of total selected features respectively. Sentinel-2 based features dominated the emergence, silking, dough, and dent stages, comprising 66%, 89%, 89%, and 100% of selected features respectively. Overall, the GNDVI feature optimization was selected more than any other vegetation index, and was especially present in the silking and dent stages where, for each growth stage, it comprised 67% of selections. As for the focal statistic window dimension optimizations, all selected Sentinel-1 based features had focal statistic window dimensions of 3x3 or greater, supporting the finding that applying such optimizations can reduce error and smooth anomalous pixel values (Luo et al., 2020). Sentinel-2-based features had varied focal statistic window dimensions across growth stages in each site-year, but in most cases trended towards smaller dimensions in the first three growth stages if selected, and larger dimensions in the final three growth stages (Table 1.7 and Figure 1.7).

These results are reflective of the unique advantages of the Sentinel-1 and Sentinel-2 satellite systems, as well as the nature of the biophysical characteristics present across maize growth stages. Changes in the physical structure of the maize crop such as leaf development and plant height primarily occur between the emergence and tassel initiation stages (Kansas State University Agricultural Experiment Station And Cooperative Extension Service, 2007). Therefore, the microwave backscatter measurements collected by Sentinel-1 based vegetation indices can provide more insight into the structural differences present across maize fields (Nasirzadehdizaji et al., 2019). Alternatively, when maize plants begin their reproductive phase starting at silking, they have achieved their maximum height and display observable changes most notably in leaf color (Kansas State University Agricultural Experiment Station And Cooperative Extension Service, 2007). These changes in color, caused by stress induced differences in chlorophyll content, can be most accurately measured by the optical sensors present in Sentinel-2 based vegetation indices (Boiarskii, 2019). This difference in vegetative and reproductive characteristics can also explain the larger focal statistic dimensions found in the earlier stages. During the early growth stages, canopy cover is sparser, allowing for differences in soil color and texture to have a greater impact on pixel values than in the later reproductive stages (Luo et al., 2020).

Table 1.7 Results of the feature selection process for each site-year.

Site-Year ID	Growth Stage	Acquisition Date	Acc. GDD Units (°C)	Vegetation Index	Focal Statistics Window Dimension	%IncMSE Value
TO_2021	Emergence	05/23/21	226.67	VV	7x7	3.47

	Tassel initiation	06/03/21	321.47	NDRE	5x5	7.27
	Tassel emergence	07/03/21	754.44	GNDVI	1x1	5.91
	Silking	07/23/21	1033.60	GNDVI	1x1	7.49
	Dough stage	08/02/21	1182.56	SAVI	1x1	6.33
	Dent stage	08/22/21	1470.82	GNDVI	1x1	12.92
RO_2022	Emergence	05/09/22	120.06	EVI	5x5	6.12
	Tassel initiation	05/30/22	333.42	VH	9x9	7.39
	Tassel emergence	06/28/22	717.62	GNDVI	7x7	8.78
	Silking	07/13/22	940.81	GNDVI	1x1	7.26
	Dough stage	08/02/22	1225.98	GNDVI	1x1	16.66
	Dent stage	08/12/22	1368.72	GNDVI	1x1	14.86
RO_2023	Emergence	05/24/23	238.62	GNDVI	9x9	7.72
	Tassel initiation	06/18/23	526.71	VV	7x7	12.99
	Tassel emergence	06/28/23	658.27	EVI	9x9	5.02
	Silking	07/23/23	989.67	GNDVI	1x1	6.56
	Dough stage	07/28/23	1070.91	EVI	9x9	5.83
	Dent stage	08/22/23	1416.87	NDVI	3x3	8.42
CO_2024	Emergence	05/22/24	97.06	VV	9x9	3.45
	Tassel initiation	06/15/24	354.46	RATIO	7x7	9.71
	Tassel emergence	07/09/24	662.34	VH	9x9	7.20
	Silking	07/21/24	821.07	VV	9x9	5.58
	Dough stage	08/19/24	1184.57	RATIO	9x9	5.66
	Dent stage	09/01/24	1351.94	NDVI	1x1	6.31
FC_2021	Emergence	05/23/21	78.94	EVI	3x3	6.18
	Tassel initiation	06/29/21	412.50	RATIO	9x9	4.83
	Tassel emergence	07/28/21	746.50	VH	9x9	3.79
	Silking	08/04/21	825.61	SAVI	3x3	2.45
	Dough stage	09/20/21	1306.33	NDRE	1x1	9.69
	Dent stage	10/18/21	1464.56	GNDVI	3x3	4.08
FC_2024	Emergence	06/01/24	175.44	RVI	9x9	2.68
	Tassel initiation	06/25/24	423.72	VV	3x3	2.92
	Tassel emergence	07/26/24	763.61	RATIO	3x3	2.38
	Silking	08/05/24	881.33	GNDVI	1x1	8.08
	Dough stage	08/30/24	1145.28	NDRE	1x1	10.62
	Dent stage	10/07/24	1479.78	GNDVI	1x1	4.26
FR_2022	Emergence	05/14/22	122.94	NDRE	5x5	5.37
	Tassel initiation	06/28/22	554.17	EVI	3x3	10.05
	Tassel emergence	07/08/22	683.06	NDVI	1x1	8.09
	Silking	08/02/22	1008.17	GNDVI	1x1	7.34
	Dough stage	08/27/22	1314.44	GNDVI	1x1	9.09
	Dent stage	09/01/22	1368.22	GNDVI	1x1	8.93
FR_2023	Emergence	05/19/23	75.83	NDVI	9x9	3.77

	Tassel initiation	07/06/23	536.56	RVI	5x5	4.26
	Tassel emergence	07/23/23	735.11	VV	5x5	3.05
	Silking	07/28/23	800.61	NDRE	3x3	7.65
	Dough stage	09/06/23	1284.61	NDVI	1x1	7.20
	Dent stage	09/16/23	1375.89	NDRE	3x3	9.54
FR_2024	Emergence	05/28/24	202.17	NDRE	5x5	6.25
	Tassel initiation	06/06/24	291.33	VV	9x9	3.88
	Tassel emergence	07/12/24	710.72	RATIO	5x5	3.88
	Silking	07/27/24	905.61	GNDVI	1x1	5.79
	Dough stage	08/11/24	1091.94	GNDVI	1x1	9.71
	Dent stage	09/10/24	1428.72	GNDVI	1x1	2.92

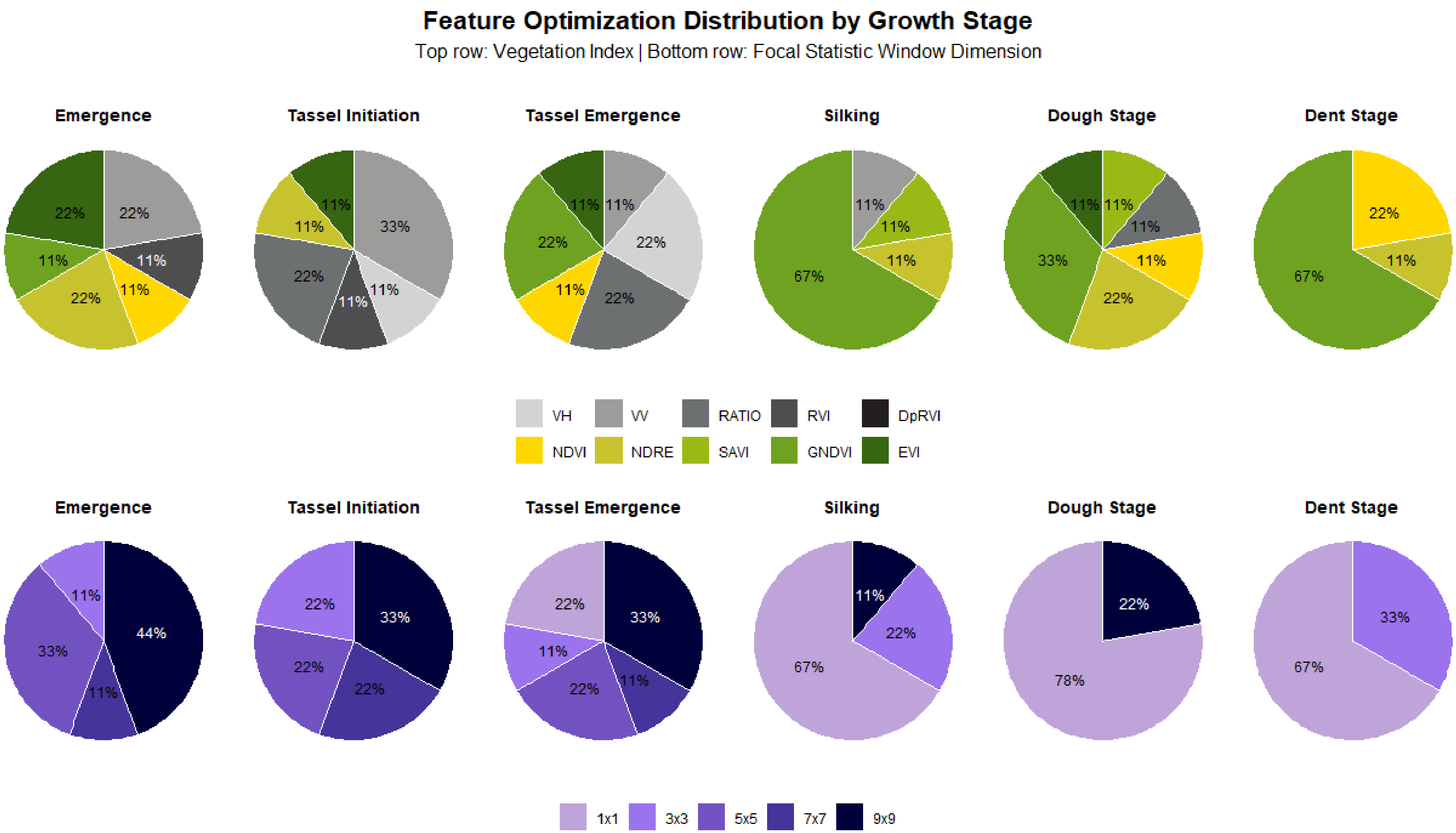


Figure 1.7 Combined site-year vegetation index and focal statistic window dimension feature optimization distributions across growth stages.

Broader analysis of the features selected for site-years within identical or relatively similar geographic locations resulted in several notable findings. The Fruita site-years, specifically 2023 and 2024, had Sentinel-2 based features selected for all but the tassel initiation and tassel emergence stages. While the Fruita 2022 site-year selected Sentinel-2 features across each growth stage, the final three growth stage feature selections were identical to the Fruita 2024 feature selections. The Fort Collins site-years were unique due to their more balanced ratio of selected features, with each site-year resulting in primarily Sentinel-1 based features for the first three growth stages, and exclusively Sentinel-2 features for the final three growth stages. The Colby site-year stands out from the rest due to its selection of Sentinel-1-based features for all but the final growth stage. The Rossville and Topeka site-years, located within only 14.5 km of each other, selected only one Sentinel-1 based feature, each within the first two growth stages.

Overall, the results of the feature selection process demonstrated the importance of integrating a diverse set of features across growth stages to account for the geospatial variability found across site-years. This process also provided insight into the variability of feature importance across growth stages within each site-year. Notably, only five of the nine site-years had a positive correlation between feature importance and increasing GDD unit accumulation (see Figure 1.8), demonstrating that features generated from images with acquisition dates closest to harvest are not always the most important features to use for generating accurate yield predictions.

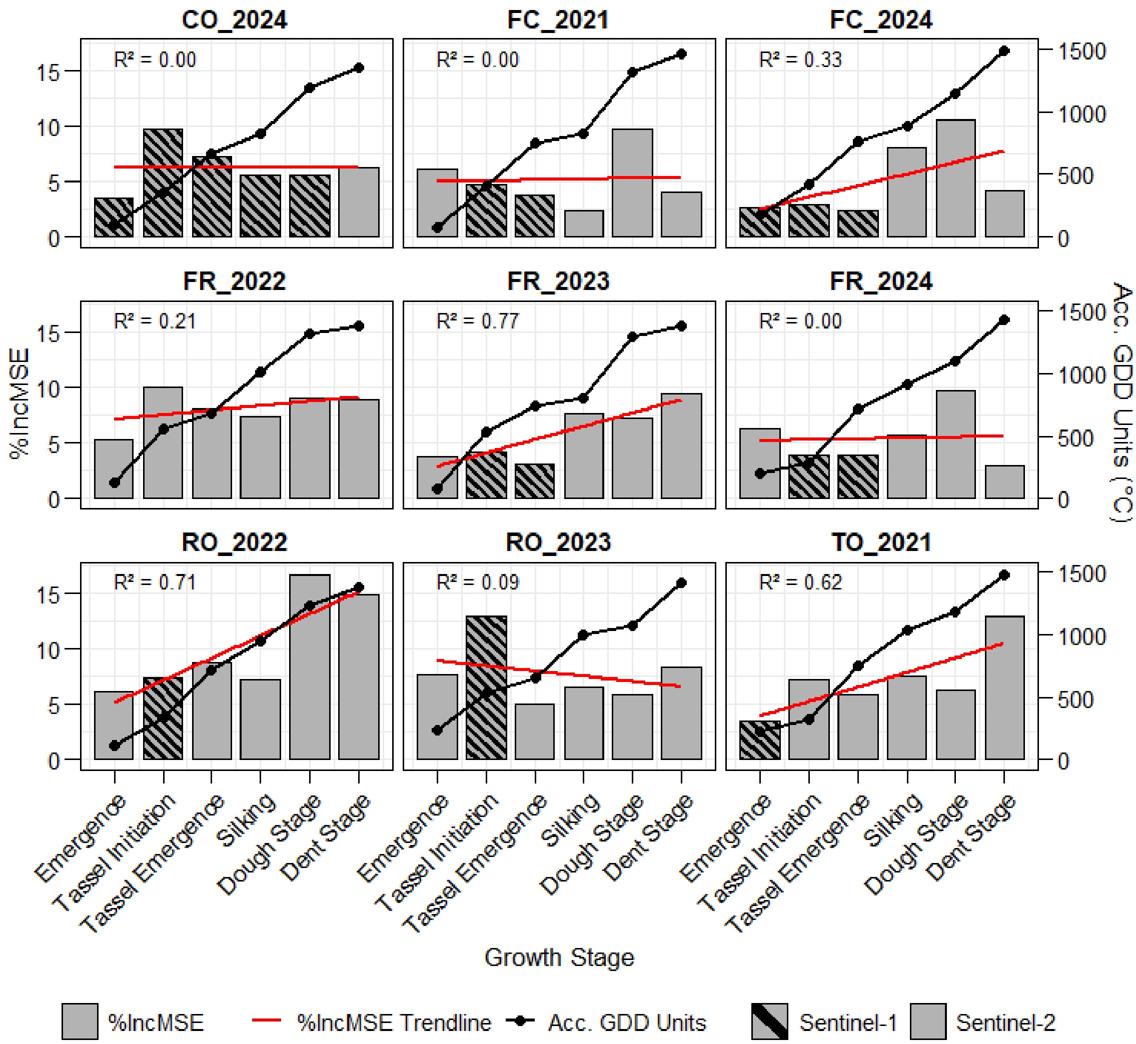


Figure 1.8 Bar plots of the Random Forest feature importance (%incMSE) value of the selected features for each GDD based growth stage across site-years, with GDD units (black line) displayed, and a trendline (red) portraying the correlation between %incMSE and accumulated GDD units.

1.3.3 Intra-Site-Year Yield Prediction Results

Across all site-years, observed maize yield varied from 1.28 to 17.75 t ha^{-1} at the pixel level (see table 1.8 below). Site-year mean maize yield ranged from 7.24 to 13.47 t ha^{-1}. Large within site-year variability occurred in the Rossville 2022, Fruita 2022 and 2023, and Topeka 2021 site-years, having pixel value yield standard deviations ranging from 2.23 to 2.83 t ha^{-1}.

Alternatively, the Fruita 2024, Fort Collins 2021 and 2024, Rossville 2023, and Colby 2024 site-years had less variation in yield, with pixel value standard deviations ranging from 0.83 to 1.29 t ha^{-1}. These differences in within-site-year variability can be attributed to the different sizes of fields used in the study ranging from 1.90 to 7.40 ha (Table 1.1). Other factors such as differences in soil fertility and texture across the AOIs may also influence the yield variability, but those factors are outside the scope of this study. Further, the notably high maximum yield value observed in the Fruita 2022 site-year (17.75 t ha^{-1}) may be a result of it being the only field under furrow irrigation.

Table 1.8 Observed yield results across site-years.

Site-Year ID	**Minimum Yield (t ha^{-1})**	**Maximum Yield (t ha^{-1})**	**Mean Yield (t ha^{-1})**	**Yield Variability Standard Deviation (t ha^{-1})**
TO_2021	5.85	13.87	10.84	2.23
RO_2022	3.30	13.54	9.49	2.83
RO_2023	10.29	15.94	13.47	1.19
CO_2024	9.77	13.29	11.61	0.83
FC_2021	6.20	12.36	9.15	1.16
FC_2024	4.52	10.85	7.94	1.29
FR_2022	1.28	17.75	13.33	2.72
FR_2023	1.78	13.07	8.91	2.46
FR_2024	3.23	11.60	7.24	1.78

The 10-Fold cross validation random forest regression models generated significantly accurate yield predictions at each site-year using the six selected features (see Figure 1.9). While most scatterplots display a generally positive linear trend, the Fruita 2022 and 2023 site-years are notable. Fruita 2022 yield predictions were significantly more accurate for higher-yielding areas, but struggled with generating accurate predictions for lower-yielding areas. Fruita 2023 had the least accurate yield predictions but still achieved an R^2 of 0.62 and an RMSE of 1.1 t ha^{-1}.

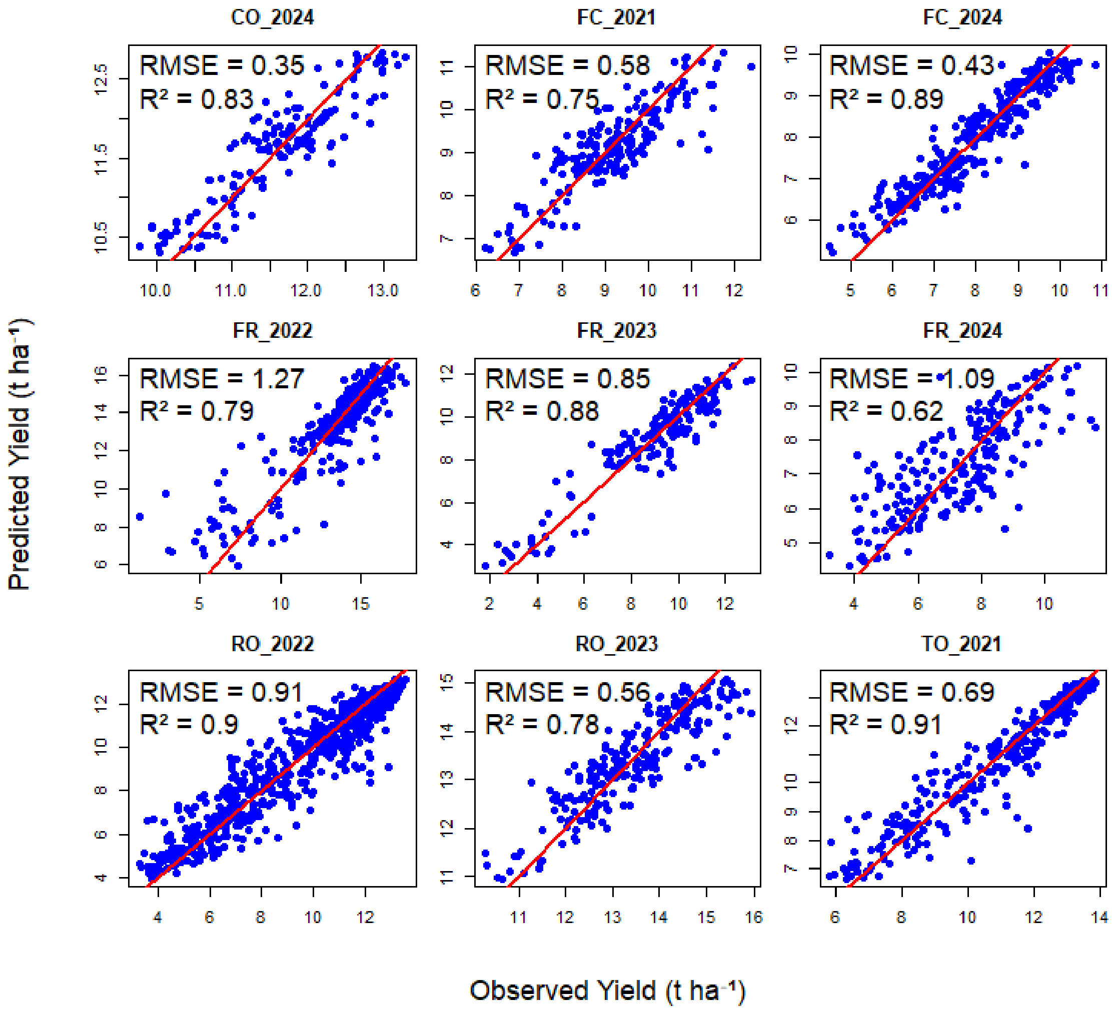


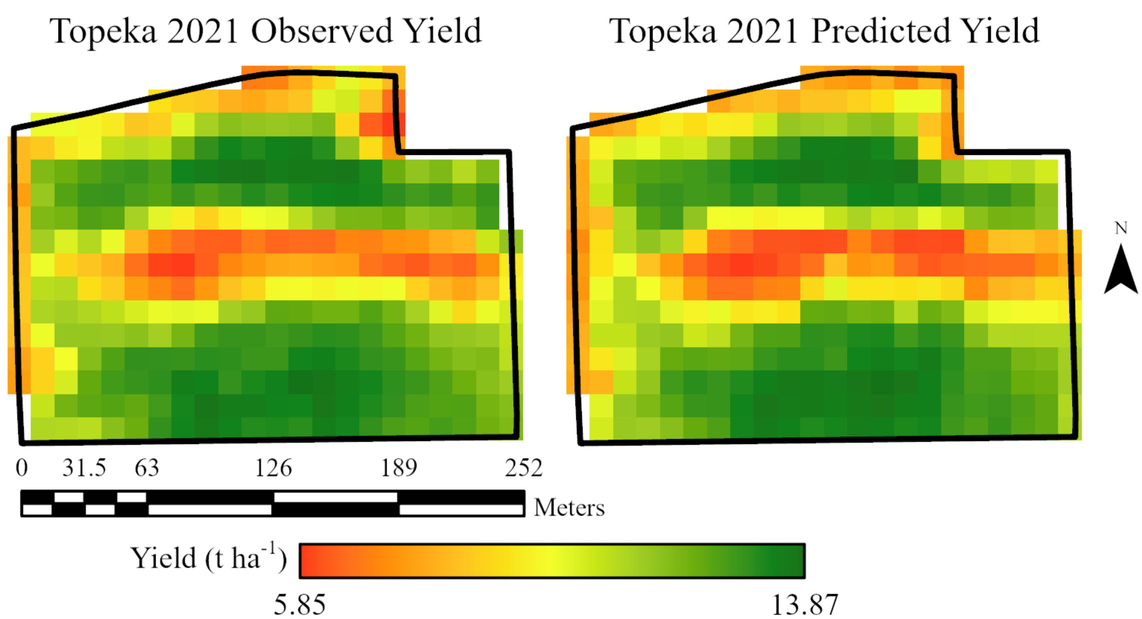
Figure 1.9 Scatterplots displaying the observed vs. predicted yield for each site-year.

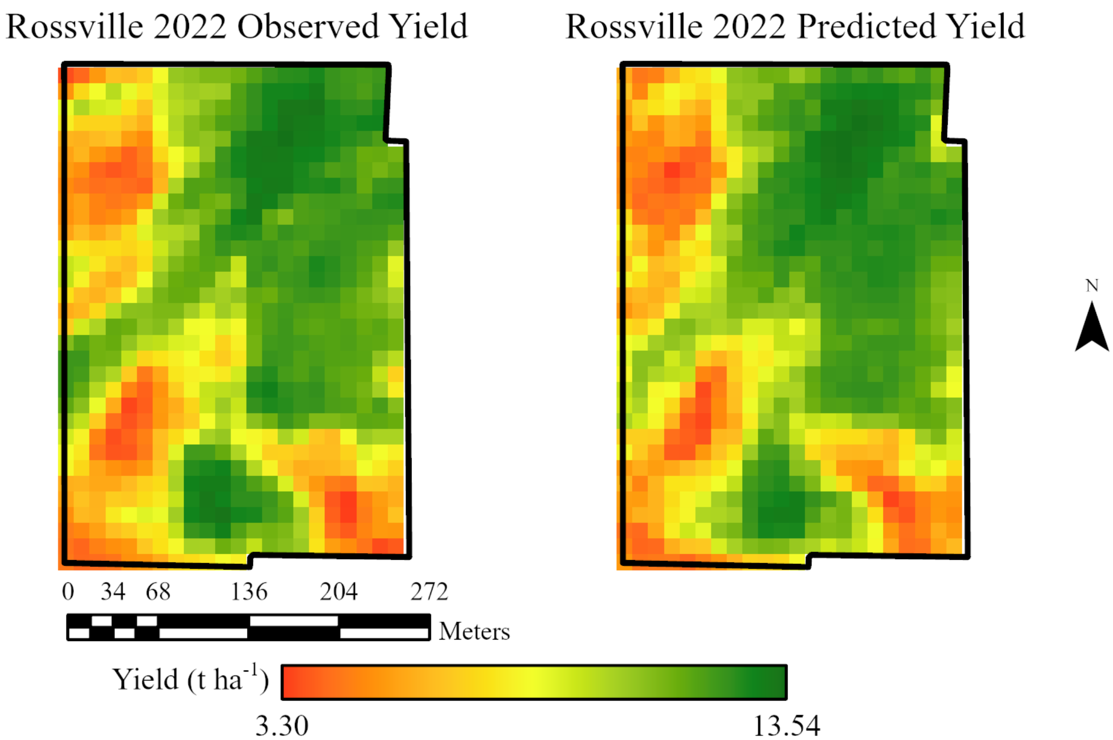
Overall, the results of the yield predictions demonstrated that this approach could generate significantly accurate yield predictions across a geographically and environmentally diverse range of site-years (see Figure 1.10 below). Additionally, it showcases the benefits of geospatially conditioning features for discrete growth stages within specific site-years.

Interestingly, the range of pixels (154 to 689) within each site-year's AOI did not significantly correlate with the accuracy of the yield predictions, demonstrating that this approach can perform well with a relatively wide range of pixels (see Table 1.9).

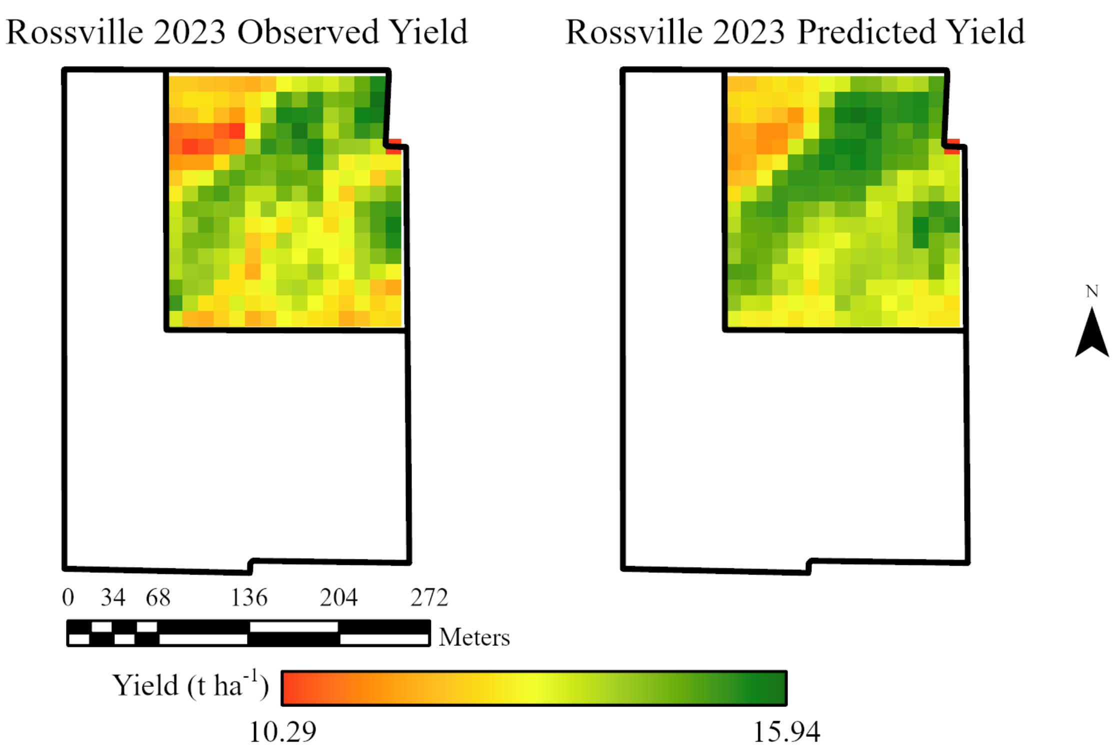
Table 1.9 Pixels within AOIs compared to RMSE and R^2 results, ordered from most to least pixels within AOI.

Site-Year ID	Pixels Within AOI	RMSE Result (t ha^{-1})	R^2 Result
RO_2022	689	0.91	0.90
FR_2022	345	1.27	0.79
FC_2024	316	0.43	0.89
TO_2021	310	0.67	0.91
RO_2023	235	0.58	0.77
FR_2024	215	1.09	0.62
FC_2021	195	0.58	0.75
CO_2024	154	0.35	0.85
FR_2023	154	0.85	0.88

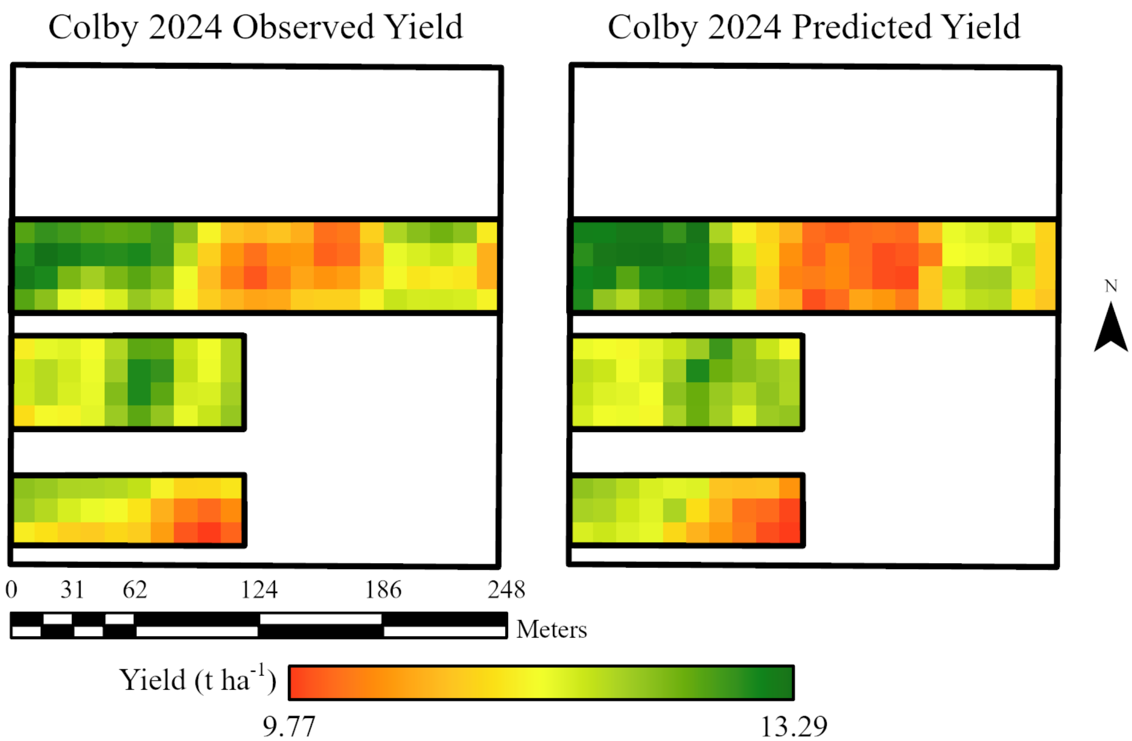


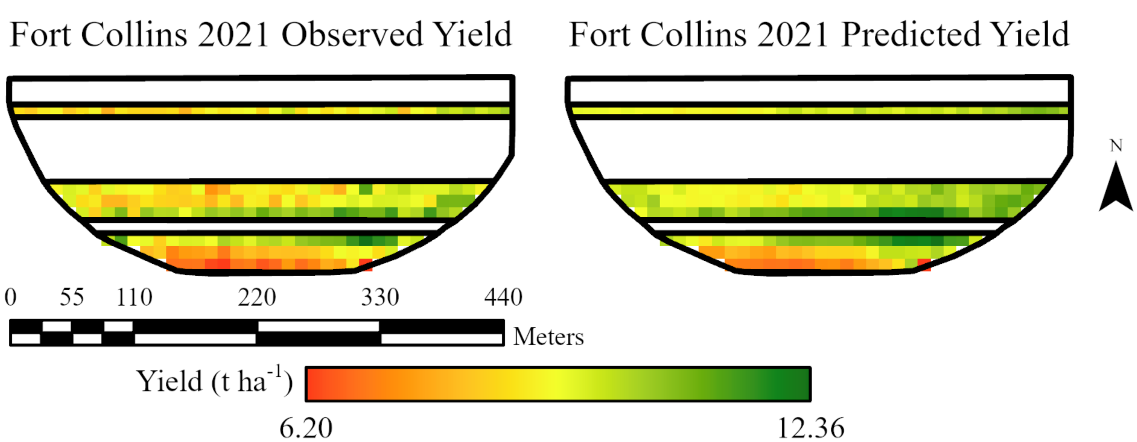


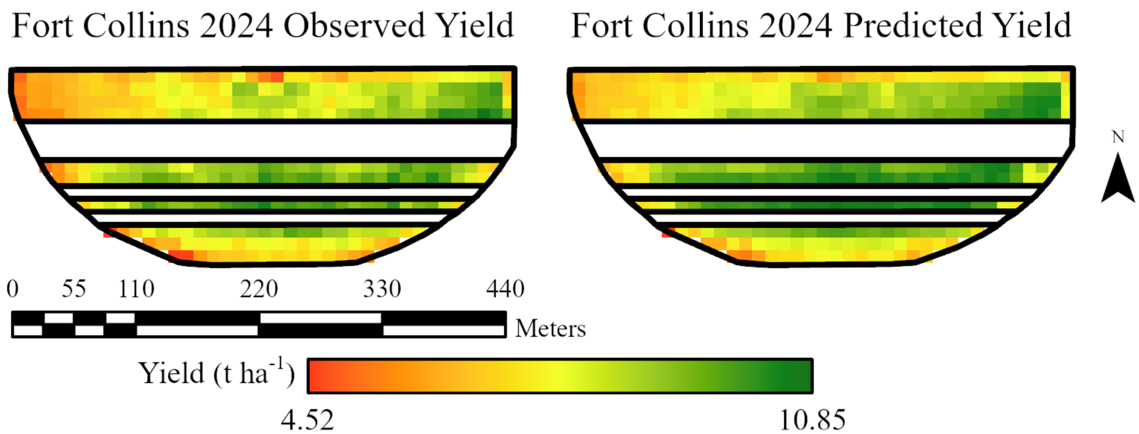
Rossville 2022 Observed Yield
Rossville 2022 Predicted Yield
N
0 34 68 136 204 272
Meters
Yield (t ha-1)
3.30
13.54



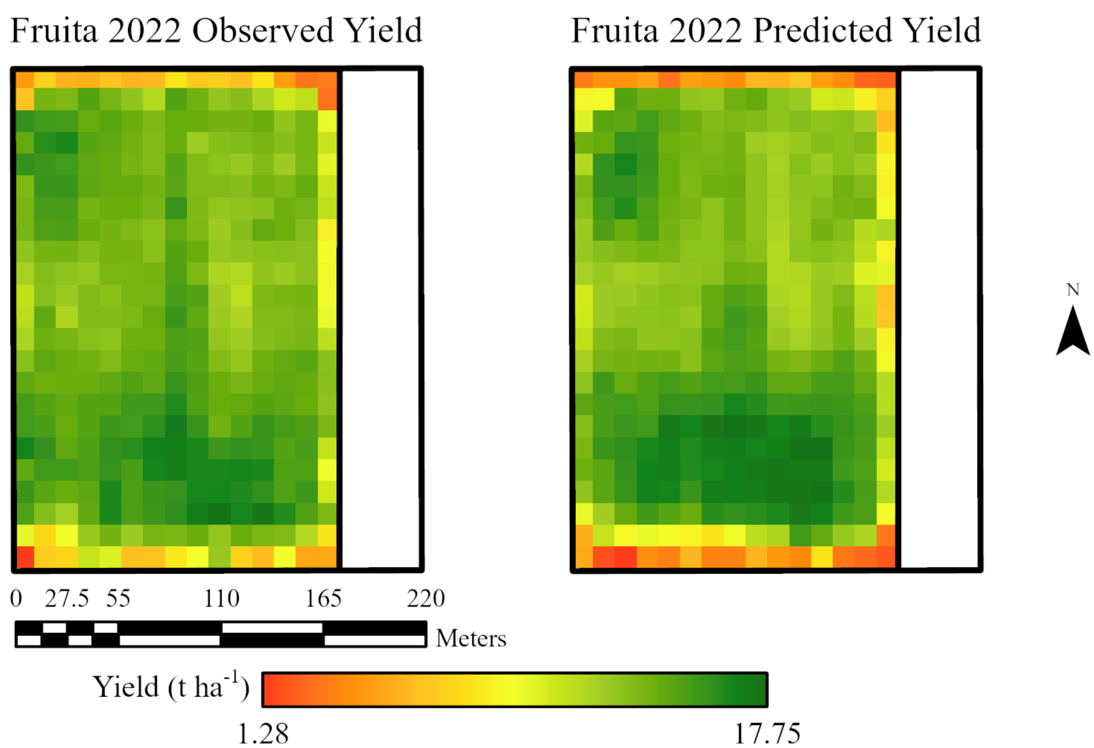
Rossville 2023 Observed Yield
Rossville 2023 Predicted Yield
N
0 34 68 136 204 272
Meters
Yield (t ha-1)
10.29
15.94


Colby 2024 Observed Yield
Colby 2024 Predicted Yield
N
0
31
62
124
186
248
Meters
Yield (t ha-1)
9.77
13.29


Fort Collins 2021 Observed Yield
Fort Collins 2021 Predicted Yield
N
0
55
110
220
330
440
Meters
Yield (t ha-1)
6.20
12.36



Fort Collins 2024 Observed Yield
Fort Collins 2024 Predicted Yield
N
0 55 110 220 330 440
Meters
Yield (t ha-1)
4.52
10.85



Fruita 2022 Observed Yield
Fruita 2022 Predicted Yield
N
0 27.5 55 110 165 220
Meters
Yield (t ha-1)
1.28
17.75

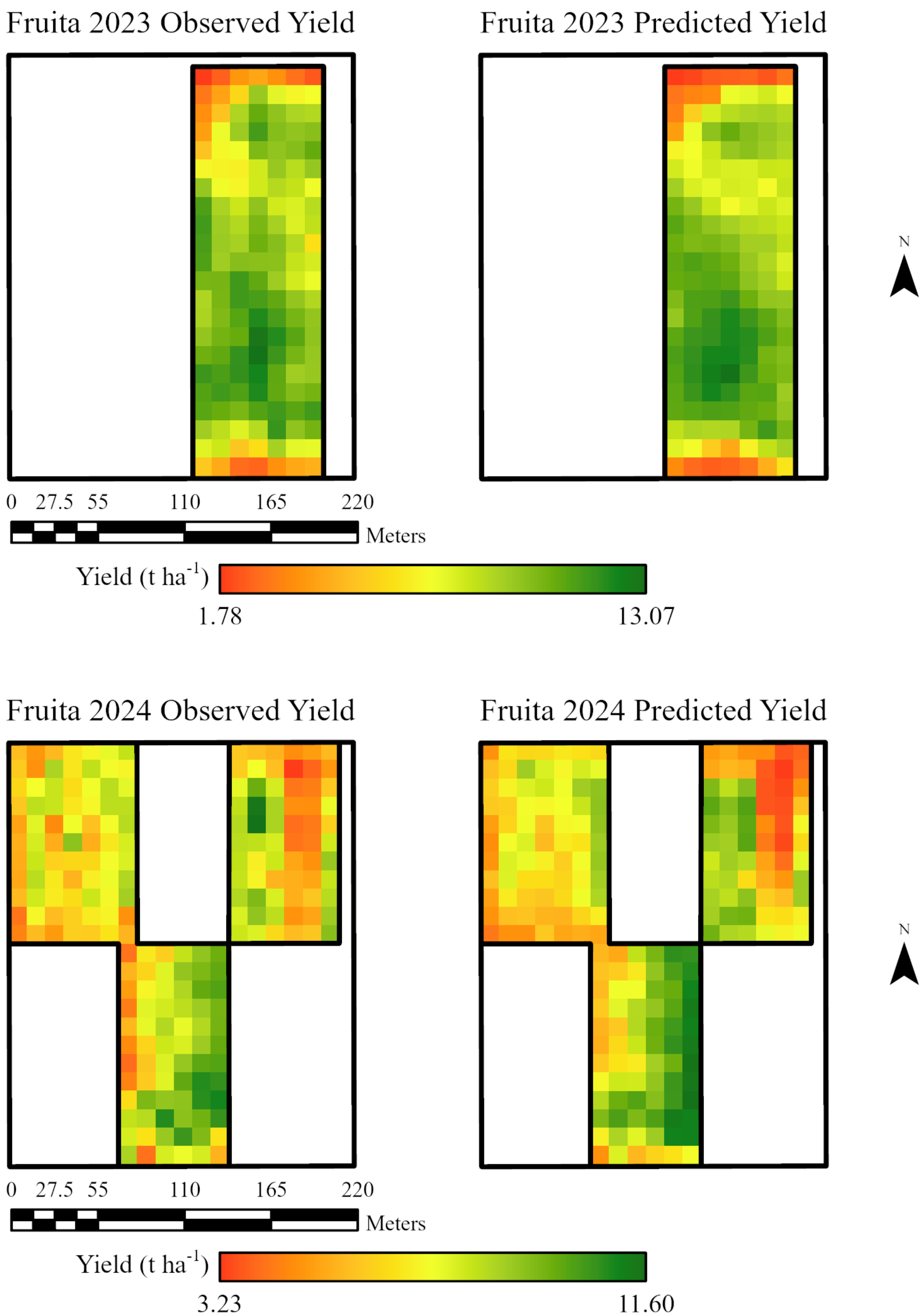


Figure 1.10 Maps depicting observed and predicted yield t ha^{-1} for each site-year. Pixel size is standardized to match that of the Sentinel-1 and Sentinel-2 satellite systems (10 m x 10 m).

1.3.4. Inter-Site-Year Yield Prediction Results

Upon completion of the inter-site-year analysis, there were no site-years that significantly benefited from using feature optimizations selected for a different site-year (see Figure 1.11). In many cases the yield results across several site-year optimizations used for a particular site-years feature set were very close in RMSE and R^2 values (see Figure 1.11). However, the results of the inter-site-year analysis demonstrate that the feature selection process succeeded at identifying significantly important feature optimizations specific to each site-year.

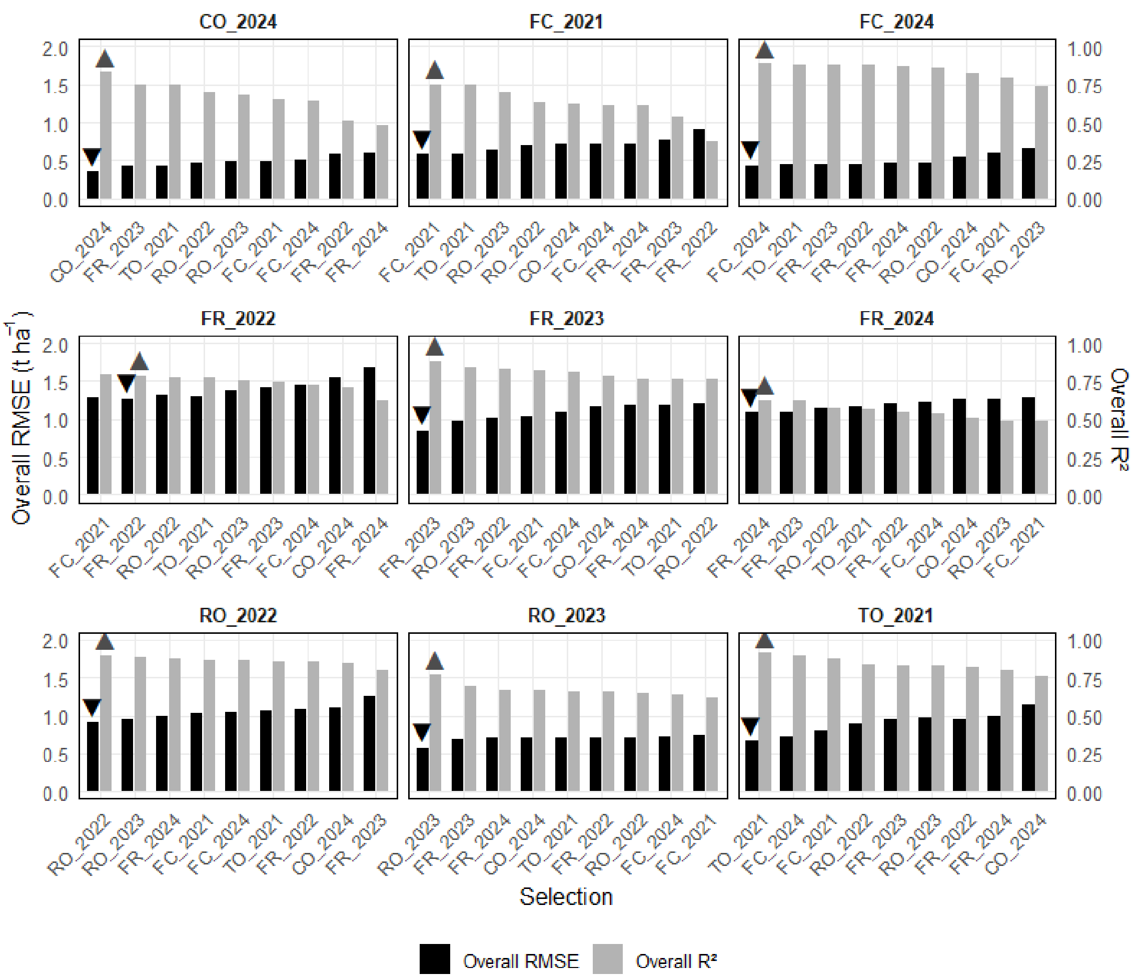


Figure 1.11. Bar charts displaying the lowest RMSE (down arrow) and highest R^2 (up arrow) results of each inter-site-year yield prediction, organized from best to worst **overall** model performance.

1.4 Conclusions

The results of this study demonstrated that by geospatially conditioning remote-sensing based model input features for discrete maize growth stages across geographically and environmentally diverse site-years, significantly accurate yield predictions can be generated (R^2 values ranged from 0.62 to 0.91 and RMSE values ranged from 0.35 to 1.27 t ha^{-1}). Further, the model input feature selection process showcased that the incorporation of a diverse set of vegetation index and focal statistic window dimension feature optimizations across discrete growth stages was preferred over a uniform feature approach. Additionally, the extra-site-year analysis demonstrated that the selected features resulted in better model performance for the site-years they were selected for. Additionally, The synergistic use of both Sentinel-1 radar and Sentinel-2 optical remote sensing datasets allowed for better insight into the spatial and temporal variability found across a diverse set of maize fields.

The implications of this study point to the advantages of accounting for the geospatial and environmental diversity present in maize fields. Future research may improve upon this study by increasing the amount of site-years, exploring different crops, and investigating the effectiveness of the approach across both irrigated and non-irrigated fields. Additionally, this study sampled a relatively small set of the available vegetation indices and focal statistic window dimensions, allowing for future studies to test even larger feature sets.

Expanding on these findings, Chapter 2 extends the scope of this study to assess the geospatially conditioned remote sensing-based models for their ability to predict maize yield at a county-scale. Additionally, USDA-NASS reported average county maize yield estimations used in conjunction with OpenET, a remote sensing-based evapotranspiration dataset, were studied to determine water productivity and the environmental characteristics and conditions that most impact it across climate divisions in Kansas and Colorado.

Chapter 2 - Remotely Sensed County-Scale Maize Yield Predictions and Identifying Climatic Characteristics of High Water Productivity

2.1 Introduction

2.1.1. Maize Production Water Productivity in Kansas and Colorado

While maize remains one of the most globally important crops, its success is directly tied to availability of water resources (Food and Agriculture Organization of the United Nations, 2025). Across the U.S. States of Kansas and Colorado, maize remains one of the most produced agricultural commodities annually, resulting in a combined 3.38 million planted hectares in 2025 (USDA-NASS, 2026). The required irrigation to sustain such production resulted in Kansas and Colorado comprising a combined 8.5% of the total U.S. irrigated farmland in a 2022 (USDA Economic Research Service, 2026). As the availability of water becomes scarcer due to declining aquifer levels, rising temperatures, and increased pressure to produce more due to an increasing population, it is imperative that water resources are used as efficiently and effectively as possible (Michon Scott, 2019; National Centers for Environmental Information, 2026; Sands et al., 2023).

Water Productivity (WP) is defined broadly as the ratio between a unit of yield output and a unit of water input (FAO, 2022). In the context of agriculture, water can be measured as Crop Evapotranspiration (ET_c), representing the sum of water returned to the atmosphere through evaporation and transpiration in a specific crop field.

$$WP = \frac{Yield}{ET_C}$$

Understanding this relationship is crucial for monitoring and improving practices in maize production for more efficient water use across Kansas and Colorado. However, calculating WP at scales large enough to make actionable decisions requires systems that can collect accurate and timely data with short turnaround analysis time of both water usage and crop production.

2.1.2. Remote Sensing-Based Maize Yield and Evapotranspiration Predictions

Predicting maize yield at a large scale has historically been difficult due to variability in land cover and environmental and climatic factors across areas larger than an individual farm field. However, the USDA-NASS provides valuable resources for collecting data necessary for conducting crop-specific yield research. Annually, the USDA-NASS publishes county-level average yield estimates based on farmer-reported and objective surveys (USDA-NASS, 2023). Additionally, the USDA-NASS provides satellite-derived Crop Sequence Boundary datasets, which allow for field-scale crop-specific analysis (Hunt et al., n.d.).

With continual advancements in satellite remote sensing, environmental metrics that previously could only be measured proximally can now be measured at larger scales and with greater frequency. Notably, remote sensing-based ET_c data collection has been successfully achieved using a myriad of satellite systems and algorithmic approaches (Fong et al., 2025). OpenET, an open source dataset, provides monthly accumulated ET_c datasets at a 30 m x 30 m pixel resolution, and has seen promising success in accurately estimating ET_c values across large geographical scales (Melton et al., 2022).

2.1.3. County-Scale Water Productivity Characteristics

Unprecedented access to the data needed to calculate WP enables deeper research into the environmental and climatic characteristics in high WP areas. Therefore, the objectives of this study were to 1) evaluate the performance of the models developed in Chapter 1 for their ability to predict yield at a county-scale, and 2) identify which climatic metrics have the highest impact on WP in maize production systems across Kansas and Colorado.

2.2 Materials and Methodology

2.2.1. Experimental Locations and Areas of Interests

All nine site-year locations from Chapter 1 were used to select four corresponding Contiguous United States (CONUS) climate divisions in Colorado and Kansas (see Table 2.1. and Figure 2.1.). From these climate divisions, 32 site-years from 17 counties were selected based on both their availability of published USDA-NASS county average yield data and county dedicated weather station data. Counties within these climate divisions historically display similar environmental characteristics (Guttman & Quayle, 1996).

Table 2.1. Site-Year Selection from Chapter 1 CONUS Climate Divisions (National Centers for Environmental Information, 2017)

Chapter 1 Site-Years	CONUS Climate Division	Chapter 2 Site-Years
FR_2022 FR_2023 FR_2024	Colorado Drainage Basin, Colorado	Delta 2022, 2023, 2024 Mesa 2022, 2024
FC_2021 FC_2024	Platte Drainage Basin, Colorado	Adams 2022, 2024 Larimer 2024 Logan 2021, 2024 Washington 2021, 2024 Weld 2021, 2024
CO_2024	Northwest, Kansas	Graham 2024 Rawlins 2024 Sherman 2024 Thomas 2024
TO_2021 RO_2022 RO_2023	East Central, Kansas	Chase 2021, 2022, 2023 Franklin 2023 Johnson 2022, 2023 Miami 2021, 2022, 2023 Osage 2022, 2023 Shawnee 2021, 2022, 2023

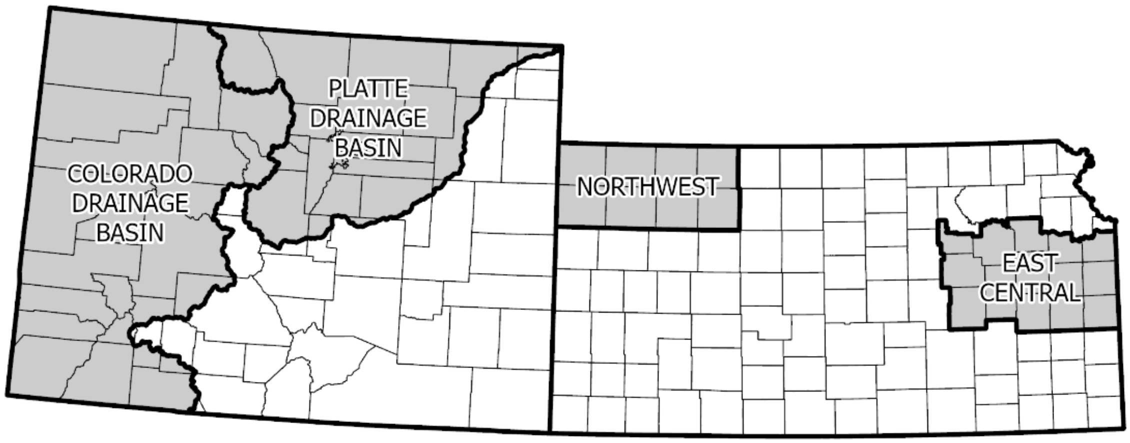
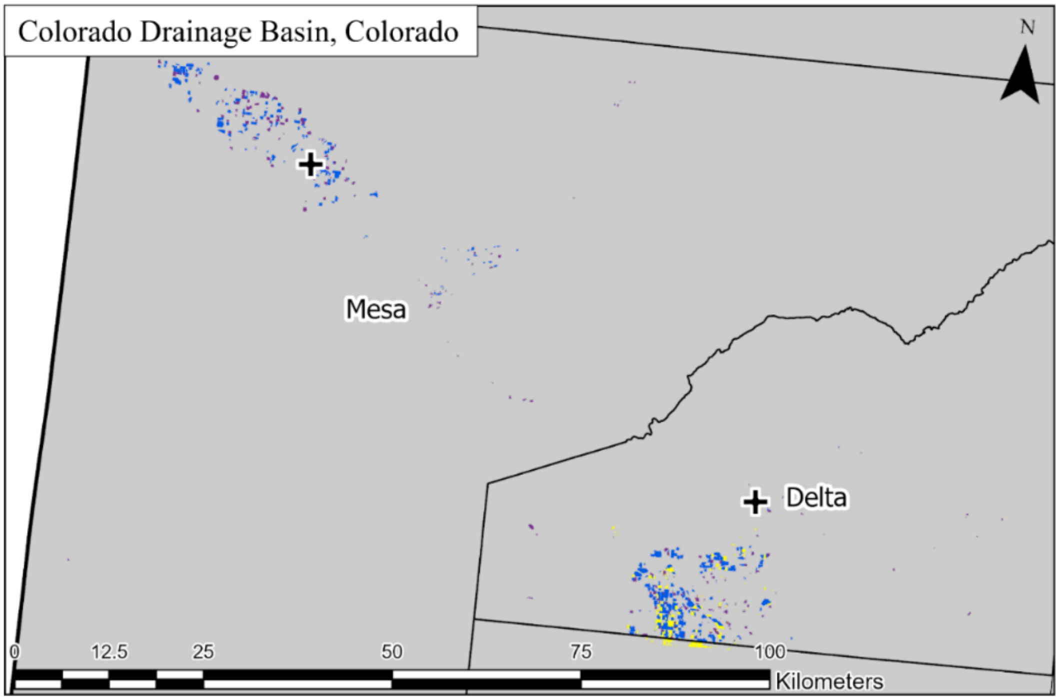
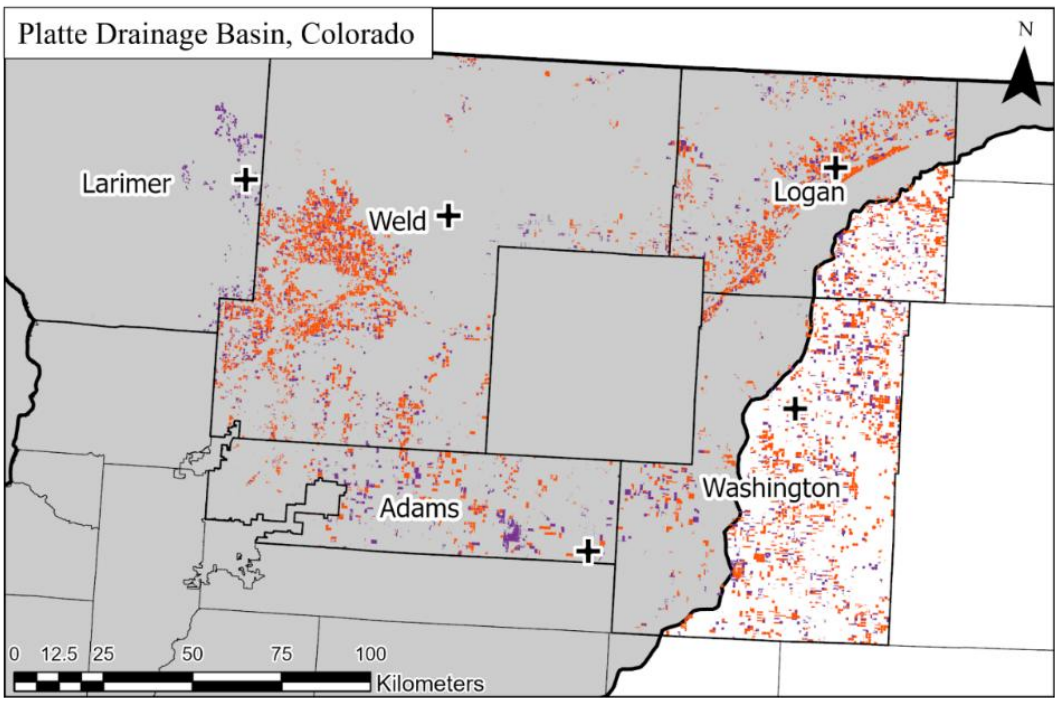


Figure 2.1 Map of CONUS Climate Divisions in Colorado and Kansas containing county-site-years.

Within each county, the USDA-NASS Crop Sequence Boundaries (CSB) dataset was used to create areas of interest (AOI) (National Centers for Environmental Information, 2017). The CSB datasets contain field-scale shapefiles of different crops derived from satellite-based Cropland Data Layers from USDA-NASS crop classification algorithms (Hunt et al., n.d.). ArcGIS Pro was used to extract maize shapefiles from the CSB datasets (see Figure 2.2.).


Colorado Drainage Basin, Colorado
N
Mesa
Delta
0
12.5
25
50
75
100
Kilometers


Platte Drainage Basin, Colorado
N
Larimer
Weld
Logan
Washington
Adams
0
12.5
25
50
75
100
Kilometers

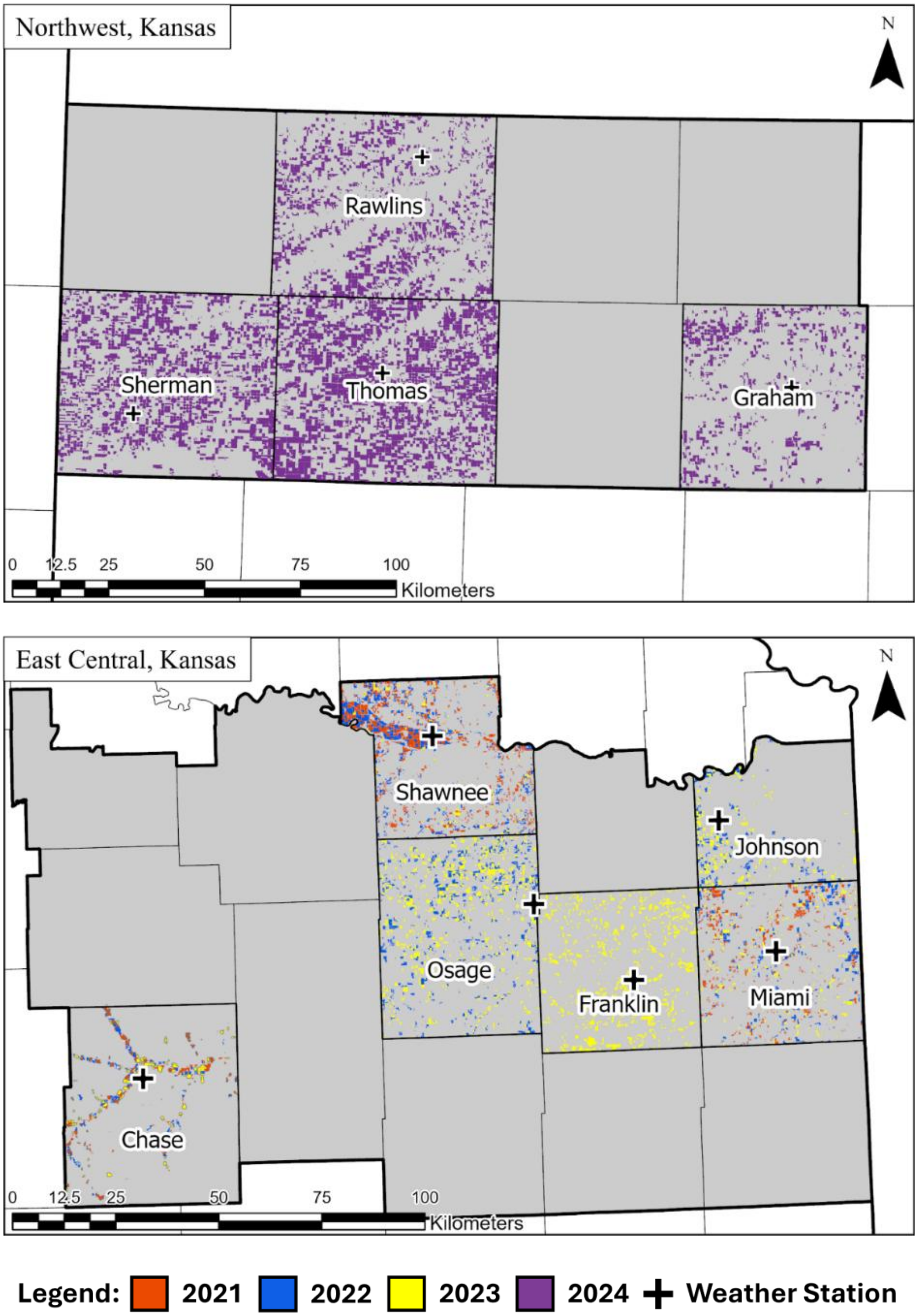


Figure 2.2 Maps of Colorado Drainage Basin, Platte Drainage Basin, Northwest Kansas, and East Central Kansas climate divisions containing 2021 (orange), 2022 (blue), 2023 (yellow), and 2024 (purple) maize Crop Sequence Boundaries. Weather stations for each county-site-year are represented by black crosses.

2.2.2. Acquisition of OpenET Data

For each CONUS climate division year, monthly OpenET Ensemble images were downloaded from Google Earth Engine (GEE) restricted to the CSB AOIs within (OpenET, 2026). These 30 m x 30 m raster datasets of monthly accumulated ET_c as an equivalent depth of water in millimeters, are derived from six well-established satellite-based ET_c mapping approaches (Melton et al., 2022). Chapter 1 site-year estimated emergence and physiological maturity dates were used to determine the months of OpenET data to download (see Tables 1.5 and 2.2.).

Table 2.2 Downloaded monthly OpenET datasets for each climate division.

Climate Division Year	Months
Colorado Drainage Basin 2022	May - September
Colorado Drainage Basin 2023	May - September
Colorado Drainage Basin 2024	May - September
Platte Drainage Basin 2021	May - October
Platte Drainage Basin 2024	May - October
Northwest Kansas 2024	May - September
East Central Kansas 2021	May - August
East Central Kansas 2022	May - August
East Central Kansas 2023	May - August

2.2.3. Generation of Sampling Points and Determination of Accumulated ETc

The ArcGIS Pro Raster Calculator tool was used to add each monthly OpenET raster dataset together, resulting in a new raster representing the accumulated ETc for each climate division year (Esri, 2026c). Next, the Raster to Point tool was employed to generate central point features for each pixel in the accumulated ETc rasters (Esri, 2025d). Finally, the Pairwise Clip tool was implemented to create new sampling point shapefiles for each site-year by extracting the Raster to Point shapefiles with each CSB AOI shapefile (Esri, 2026b). These new sampling point shapefiles were then used to sample the accumulated ETc rasters, export the data to parquet table files due to their large size, and calculate the average accumulated ETc for each CSB site-year with an R Studio script (see Figure 2.3.)

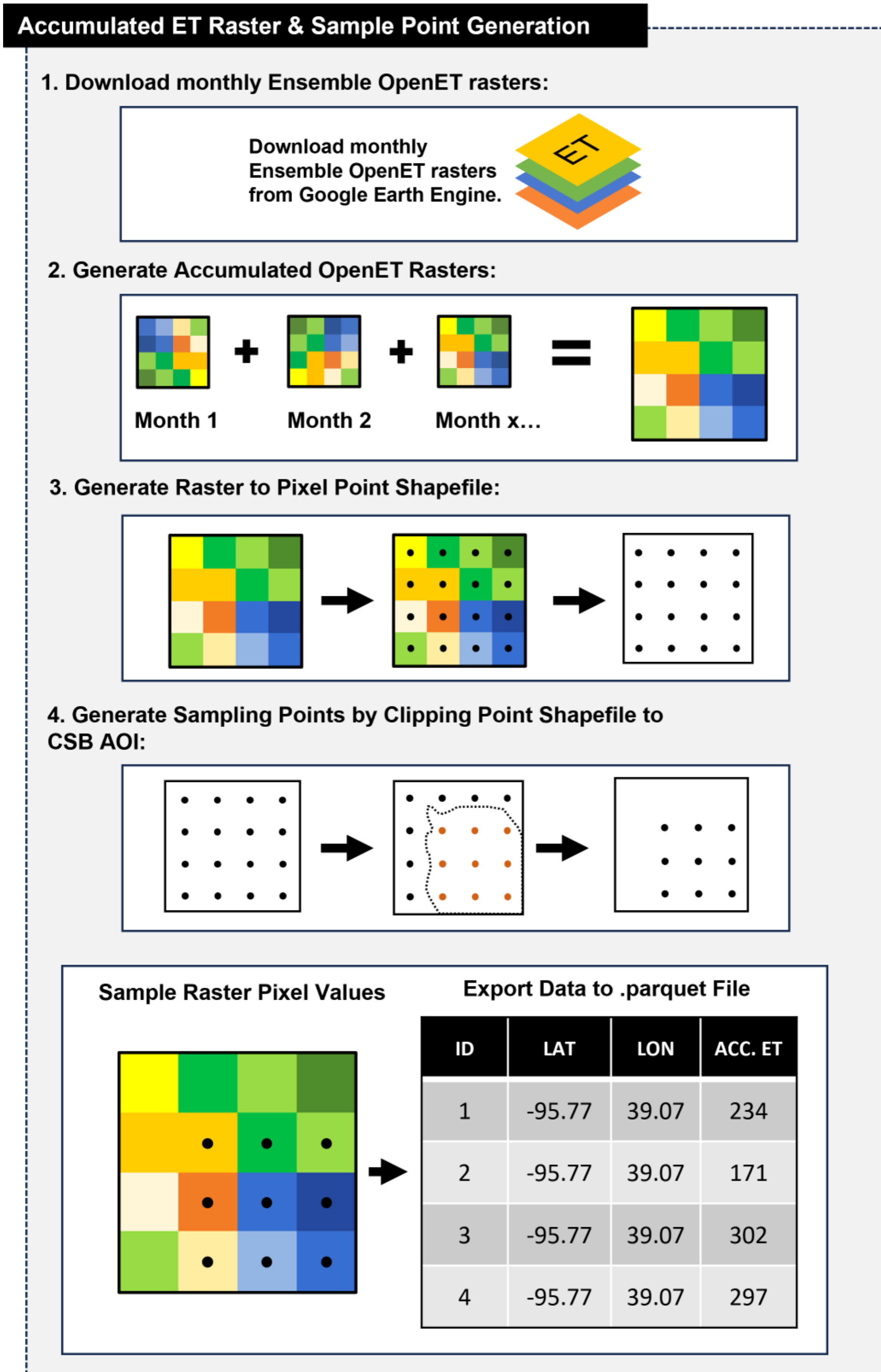


Figure 2.3 Diagram showing the process of downloading OpenET imagery, calculating accumulated OpenET, and extracting pixel data.

2.2.4. Determination of Satellite Imagery to Download

Sentinel-1 and Sentinel-2 imagery with full climate division year CSB AOI coverage were downloaded from GEE using the same date ±3 days as the corresponding Chapter 1 site-years (see table 1.8). If imagery was not available within the ±3 day window to create the original selected feature, the next available best ranked feature was used (see section 1.2.8).

2.2.5. Yield Prediction Feature Generation and Data Extraction

For each climate division year, the Mosaic to New Raster tool in ArcGIS Pro was used to select and combine the downloaded Sentinel-1 or Sentinel-2 images into new mosaic raster files for the six growth stages defined in Chapter 1 (Esri, 2026a). When working with satellite imagery, portions of images often overlap. For any overlapping pixels in this study, the average pixel value was calculated during mosaicking. Additionally, all Sentinel-1 images first had to be converted from their standard dB unit scale to Linear scale before using the Mosaic to New Raster tool. Next, the same ArcPy scripts from Chapter 1 were used to apply the VI and focal statistic window optimizations to each mosaic. Finally, a new batch-sampling ArcPy script was implemented to sample the mosaic features using the previously generated sampling point shapefiles for each CSB AOI. The sampled data was then exported to parquet table files for later use in the yield prediction section (see Figure 2.4. below).

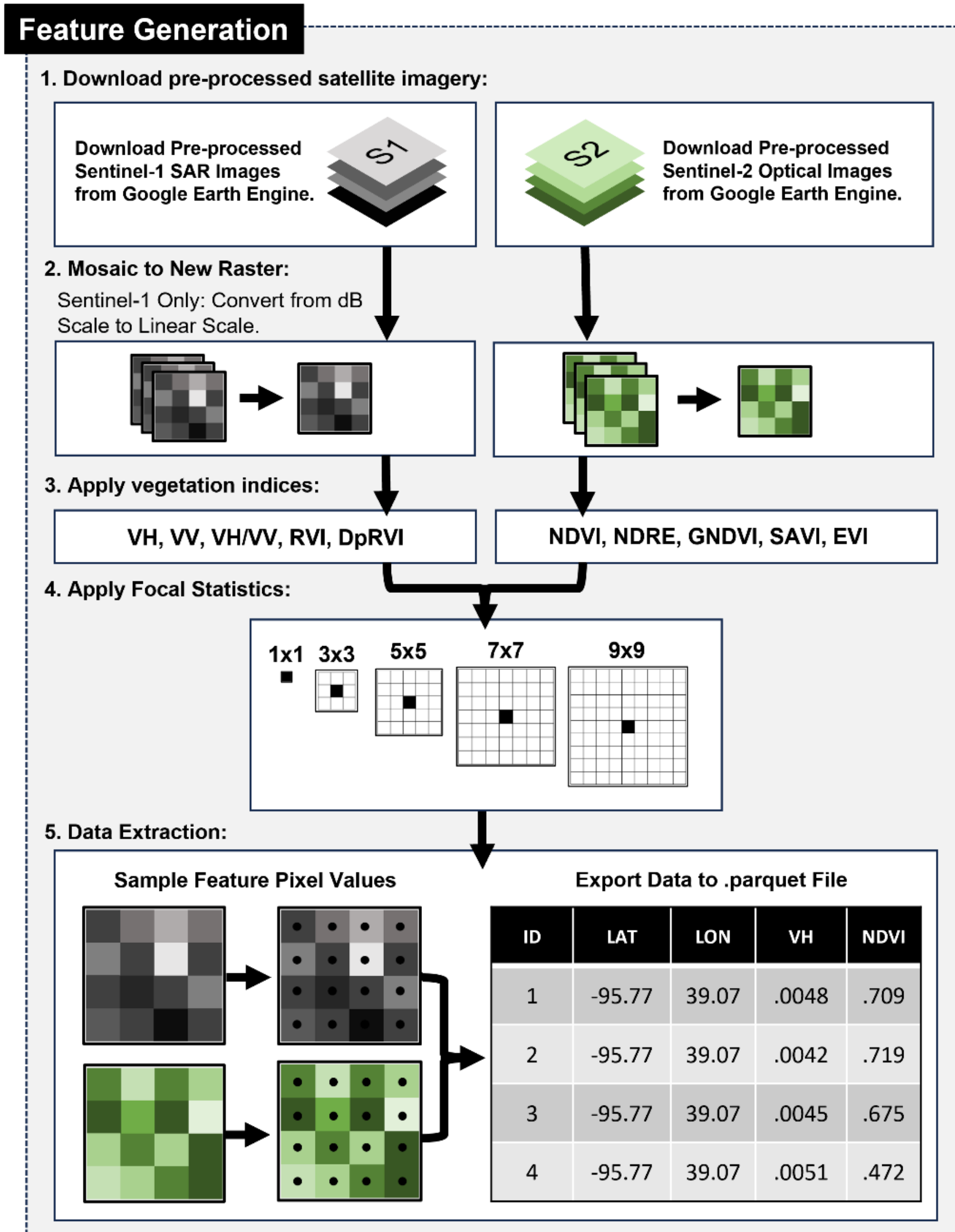


Figure 2.4 Diagram showing the process of downloading Senintel-1 and Sentinel-2 satellite imagery, creating mosaics, generating features by applying VI and focal statistic optimizations, and extracting data from pixels.

2.2.6. Determination of County Average Maize Yield

Government published county average maize yield data were collected and used as the baseline for the final county average maize yield predictions (USDA-NASS, 2026). The USDA-NASS county average maize yield estimations rely on voluntary grower-reported and objective surveys (USDA-NASS, 2023).

2.2.7. Generation and Analysis of County Average Maize Yield Predictions

Training data for county-scale yield predictions were generated from the Chapter 1 yield data and corresponding sampled feature sets. If different features were used in any of the climate division datasets because the original feature selection data were unavailable, the training data was modified to include the corresponding feature selection data from Chapter 1.

These training datasets were then loaded into an R Studio script and used to train random forest regression models for each site-year. The trained models were then used to predict maize yield for the sampled county-scale feature data stored in the parquet tables generated in section 2.2.5. Finally, the average predicted maize yield for each site-year was calculated and compared to the USDA-NASS reported yield predictions. Next, RMSE and R^2 values were generated to determine the overall accuracy of the 32 yield predictions compared to the reported USDA-NASS average yields. Additionally, line plots were generated for each climate division to investigate the ability of the prediction models to track accurate trends in high and low yields.

2.2.8. Acquisition of Climatic Data Sets

Ten environmental metrics were selected for use in this study to determine characteristics that correlated best with water productivity based on available matching data from Kansas Mesonet and CoAgMET weather stations (Colorado Agriculture and Meteorological Network, 2026; Kansas Mesonet, 2026). Additionally, 1/3 Arc Second Digital Elevation Models (DEM) and gridded soil classification raster files were downloaded for each climate division year (Soil Survey Staff, 2026; United States Geological Survey, 2026). These raster files were then sampled in the same manner as the OpenET datasets for each site-year CSB AOI to calculate average elevation above sea level and soil classification variability (determined as the standard deviation of the percentages of sampled soil types). The resulting data in Table 2.3. summarizes the data collected for each of the 32 site-years.

Table 2.3 Environmental data collected for each site-year and their sources.

Characteristics	Units	Source
Average Daily Maximum Air Temperature	(C°)	Colorado Agriculture and Meteorological Network, 2026; Kansas Mesonet, 2026
Average Daily Minimum Air Temperature	(C°)	
Average Daily Relative Humidity	(%)	
Accumulated Precipitation	(mm)	
Average Daily Wind Speed	(km day^{-1})	
Average Daily Maximum Soil Temperature at 5 cm	(C°)	
Average Daily Maximum Soil Temperature at 5 cm	(C°)	
Accumulated Daily Solar Radiation	(MJ m^{-2})	
Average Elevation Above Sea Level	(m asl)	United States Geological Survey, 2026
Soil Classification Variability	(% SD)	Soil Survey Staff, 2026

2.2.9. Correlation Between Climatic Characteristics and Water Productivity

Water productivity was calculated by dividing the USDA-NASS reported average county yields by the accumulated county-average ET_{OpenET} calculated in section 2.2.3 based on the corresponding Chapter 1 site-years planting and harvest dates.

$$Average\ County\ WP\ (kg\ mm^{-1}) = \frac{USDA\ NASS\ Average\ County\ Maize\ Yield\ (kg\ ha^{-1})}{Average\ Accumulated\ ET_{OpenET}\ (mm)}$$

The resulting kilograms of yield per hectare per millimeter of water (kg mm^{-1}) values were subsequently analyzed for correlation to the environmental metric data. A bar chart was generated to display the water productivity for each individual site-year to identify the site-years and corresponding climate divisions with the highest water productivity values. Next, to better understand the variability in environmental characteristics of each climate division, a Radar Plot was generated to display average values of each metric for each of the four climate divisions. Then, a Pearson Correlation Matrix was generated to visualize the overall linear relationships between each of the ten climatic and environmental metrics and water productivity. And finally, a random forest regression model was run to generate predictive %incMSE feature importance values for each environmental metric to better understand non-linear relationships with water productivity. By exploring these three methods, the most important environmental metrics were identified.

2.3 Results and Discussion

2.3.1. County-Scale Maize Yield Prediction Results

USDA-NASS reported county average yield ranged from 2.73 to 15.82 t ha^{-1} across all site-years in the study. Within climate divisions, average yields ranged from 12.66 to 15.82 t ha^{-1} for the Colorado Drainage Basin, 2.73 to 13.34 t ha^{-1} for the Platte Drainage Basin, 4.79 to 7.59 t ha^{-1} for Northwest Kansas, and 5.07 to 9.66 t ha^{-1} for East Central Kansas. Variability in USDA-NASS reported average county yields remained relatively stable (standard deviations ranging from 1.24 to 1.46 t ha^{-1}) for all but the Platte Drainage Basin, which had a standard deviation of 4.06 t ha^{-1}. The yield predictions resulting from the scaled up Chapter 1 models generally overpredicted and underpredicted average county yields compared to those reported by the USDA-NASS (see Figure 2.5). The weak model performance can be attributed to several factors.

Primarily, the limited training data used for each site-year restricted the diverse pixel values of each satellite image to that of one field. Additionally, the Chapter 1 yield predictions were made for fields that were irrigated, while the CSB AOIs in this chapter contained both irrigated and non-irrigated fields. Further, the features generated for each site-year were derived from specific Chapter 1 site-year planting dates and growth stage development windows, which does not account for the variation in planting dates present across an entire county. However, despite the predicted average yield results being overall inaccurate, when sorting results into individual climate divisions, the prediction models captured trends in yield variability moderately well (Figure 2.6). Notably, the Colorado Drainage Basin and East Central Kansas climate divisions displayed similar trends in high and low average yields.

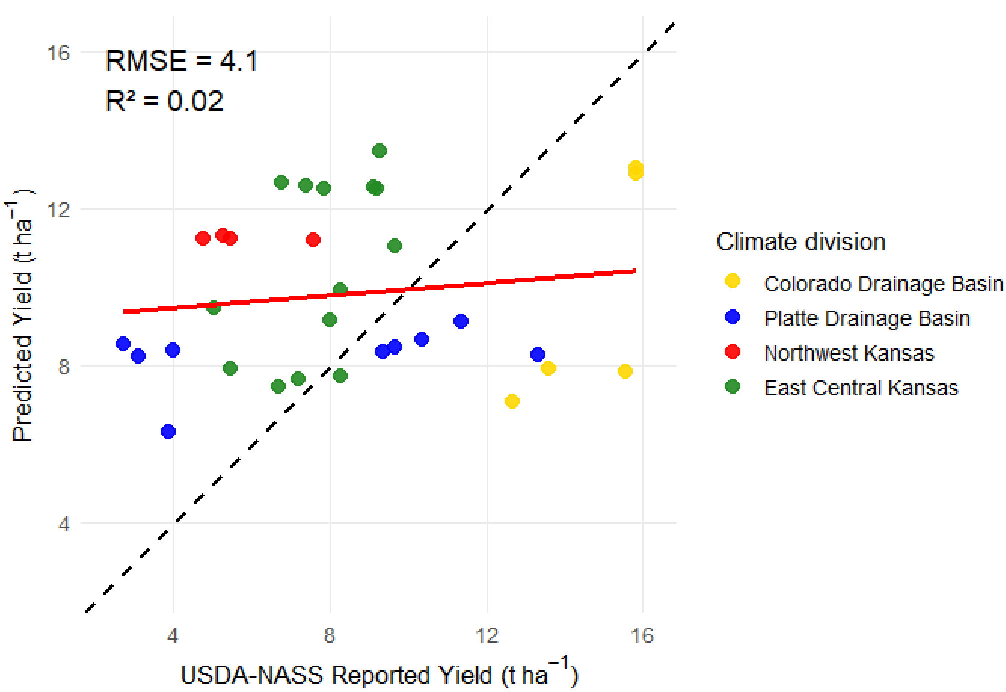
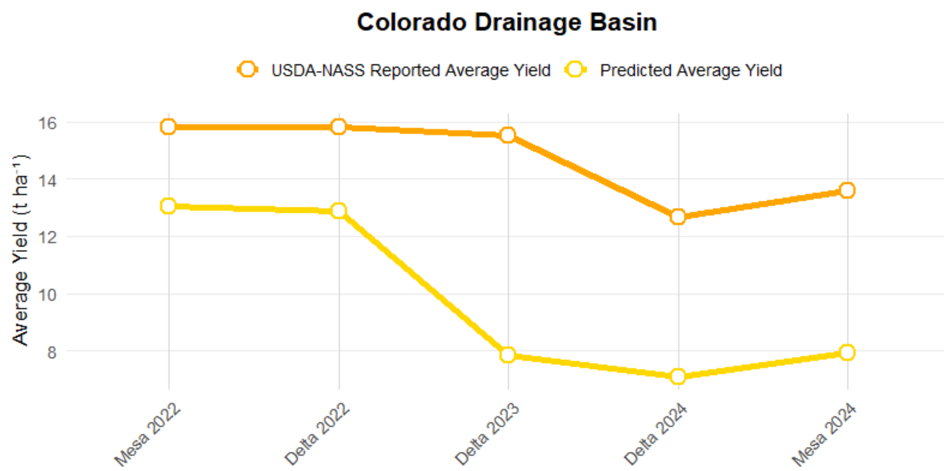
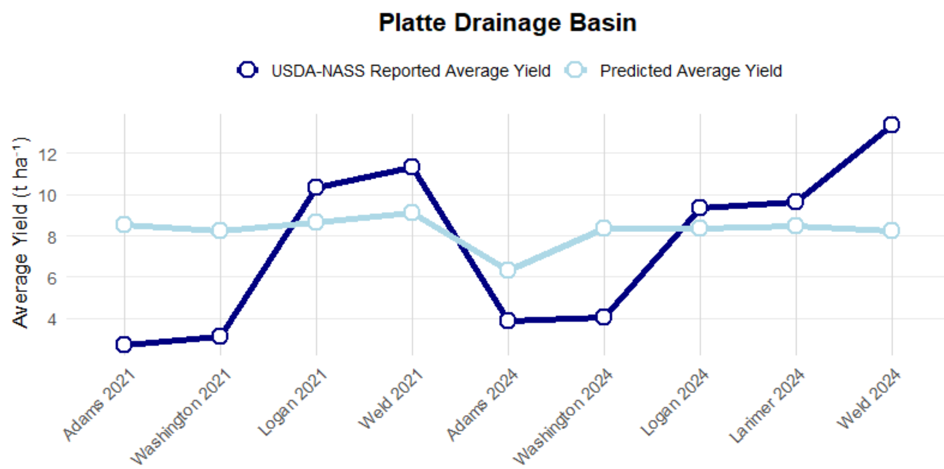
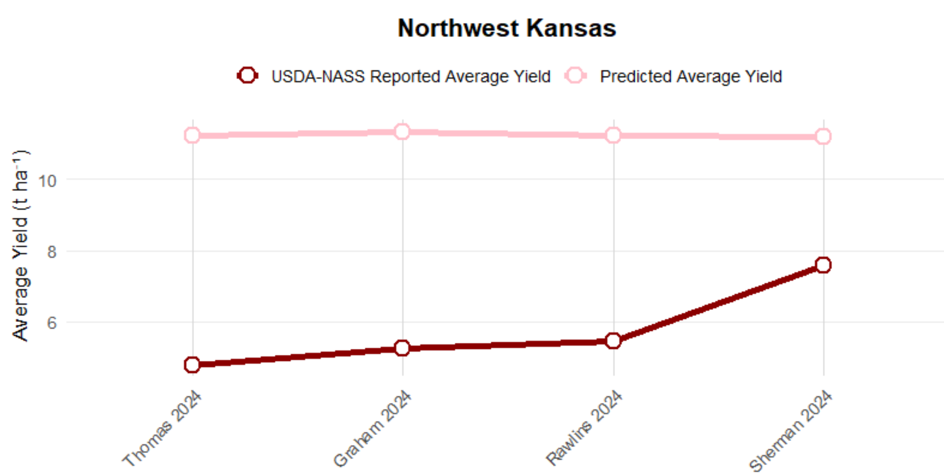


Figure 2.5 Scatterplot depicting the overall performance of the county scale average maize yield predictions versus the USDA-NASS reported average yield, colored by climate division.


Colorado Drainage Basin
USDA-NASS Reported Average Yield
Predicted Average Yield
Average Yield (t ha⁻¹)
16
14
12
10
8
Mesa 2022
Delta 2022
Delta 2023
Delta 2024
Mesa 2024


Platte Drainage Basin
USDA-NASS Reported Average Yield
Predicted Average Yield
Average Yield (t ha⁻¹)
12
10
8
6
4
Adams 2021
Washington 2021
Logan 2021
Weld 2021
Adams 2024
Washington 2024
Logan 2024
Larimer 2024
Weld 2024


Northwest Kansas
USDA-NASS Reported Average Yield
Predicted Average Yield
Average Yield (t ha⁻¹)
10
8
6
Thomas 2024
Graham 2024
Rawlins 2024
Sherman 2024

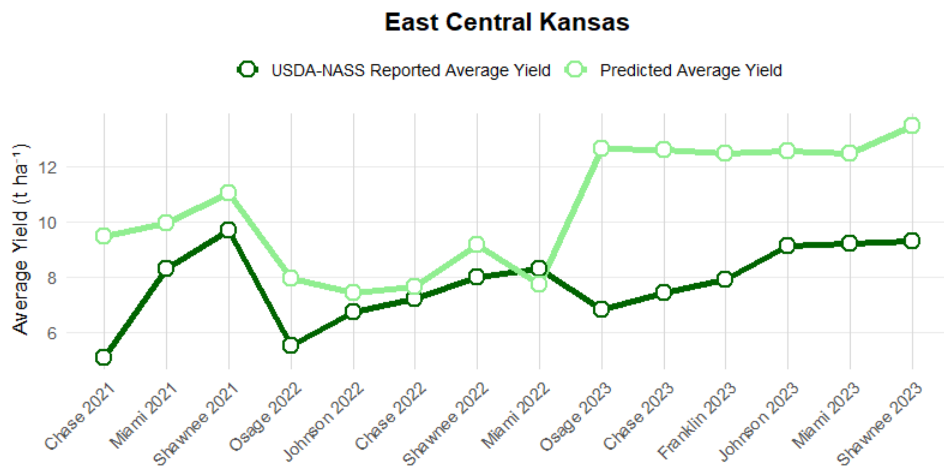


Figure 2.6 Line plots displaying USDA-NASS reported and predicted average county yields for site-years within each climate division.

2.3.2. Water Productivity Results and Identification of Impactful Environmental Characteristics

Water productivity values across all site-years ranged from 0.72 to 3.22 kg mm^{-1} (Figure 2.7). The site-years with the highest water productivity were located within the Colorado Drainage Basin climate division, with values ranging from 2.45 to 3.22 kg mm^{-1}, and an overall average water productivity value of 2.85 kg mm^{-1}. The Platte Drainage Basin and East Central Kansas climate divisions displayed moderate water productivity, with values ranging from 0.72 to 2.41 kg mm^{-1} and overall climate division averages of 1.45 and 1.61 kg mm^{-1}, respectively. The Northwest Kansas climate divisions reported the lowest overall water productivity with values ranging from 0.97 to 1.41 kg mm^{-1}, and an overall climate division average of 1.11 kg mm^{-1}. High variability exists across climate divisions, indicating a complex relationship between environmental characteristics and water productivity (Figure 2.8).

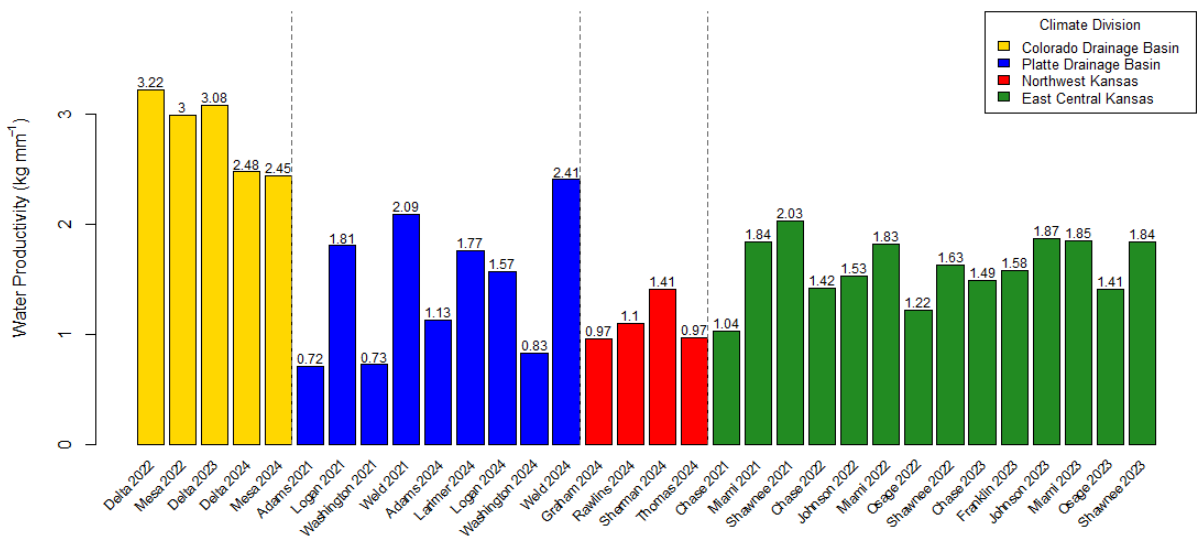


Figure 2.7 Bar chart displaying the OpenET and USDA-NASS based water productivity (kg mm^{-1}) by site-year, organized by climate division.

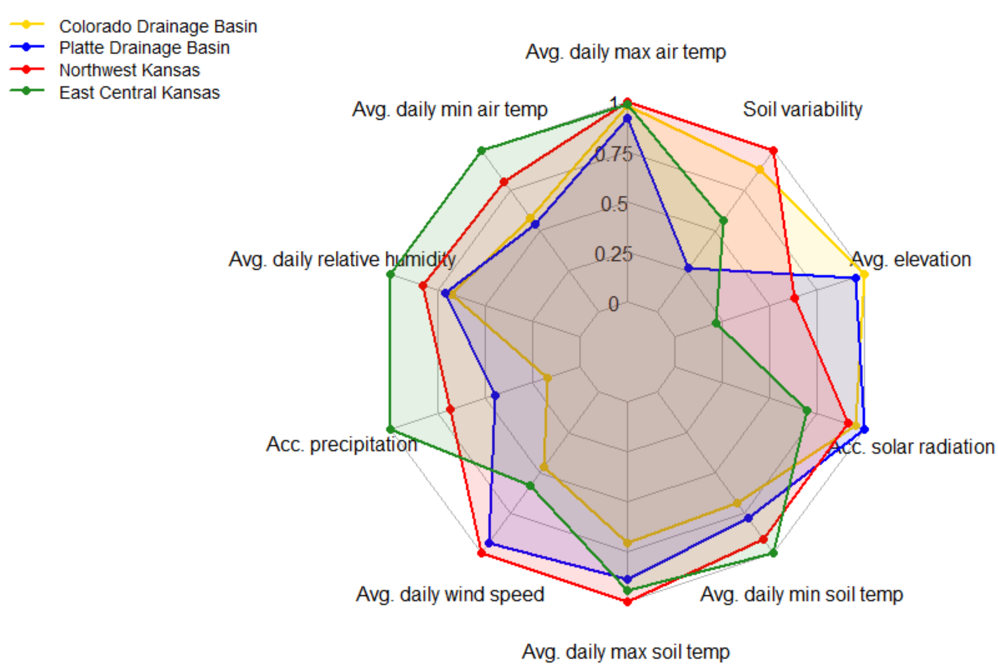


Figure 2.8 Radar plot depicting the environmental profiles of each climate division based on averages of each environmental factor value, normalized from 0-1.

The below Pearson Correlation Matrix displays the linear relationships between each of the ten climatic and environmental metrics with water productivity. Indicating agreement with the radar plot in Figure 2.8, average daily wind speed had a significant negative relationship with water productivity (r = -0.68). Additionally, average daily minimum and maximum soil temperatures, and accumulated precipitation showed moderately significant negative relationships with water productivity (r = -0.45, -0.41, and -0.35, respectively).

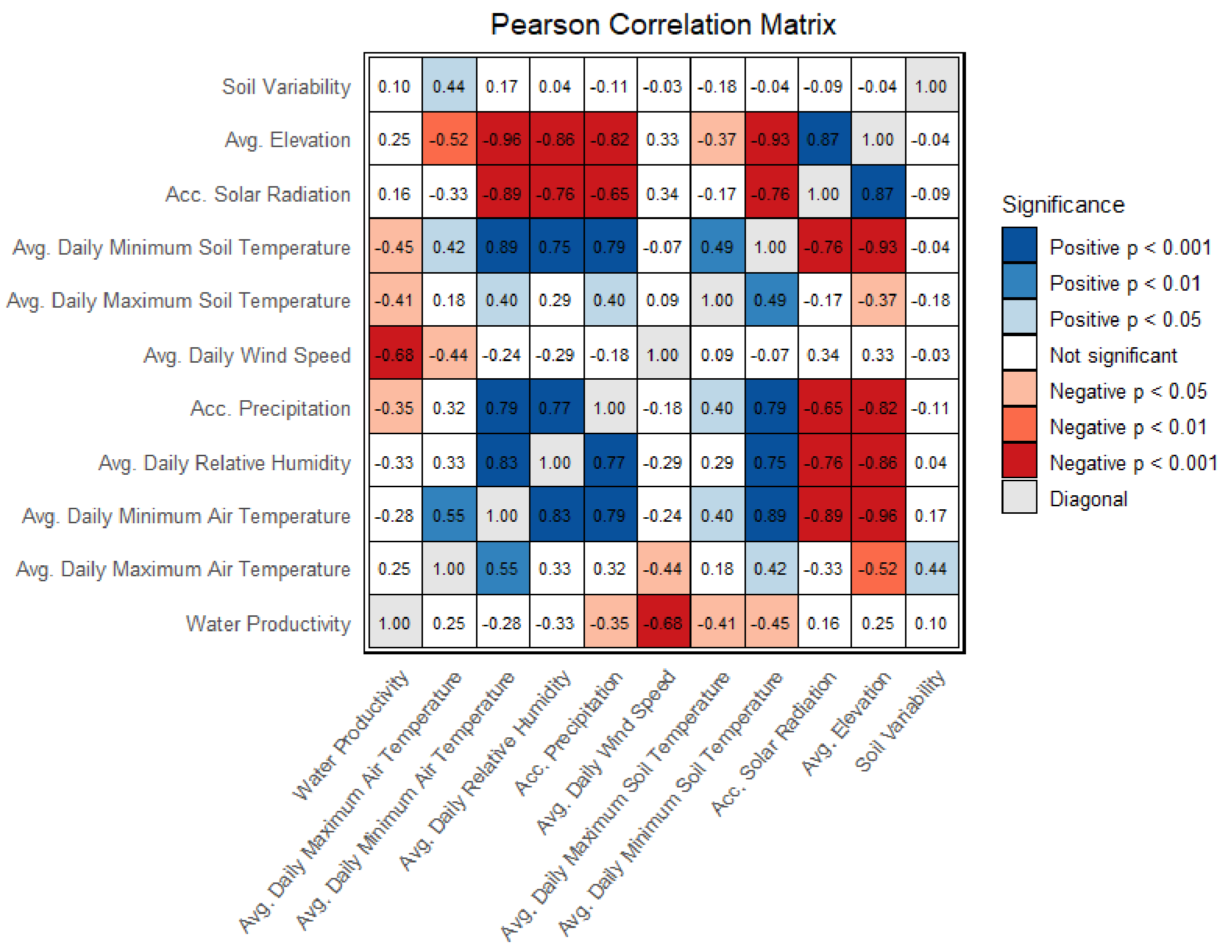


Figure 2.9 Pearson correlation matrix of water productivity and environmental factors across all 32 site-years, with negative and positive linear correlation significance shaded in red and blue, respectively.

Further, %incMSE feature importance scores generated from a random forest regression model also identified average daily wind speed as the most impactful environmental predictor of water productivity, resulting in a %incMSE value of 12.84 (Figure 2.10). Also identified as moderately impactful environmental factors, with %incMSE scores >5, included average daily relative humidity (9.19), accumulated precipitation (7.36), average elevation (6.10), average daily minimum soil temperature (5.67), and soil variability (5.10).

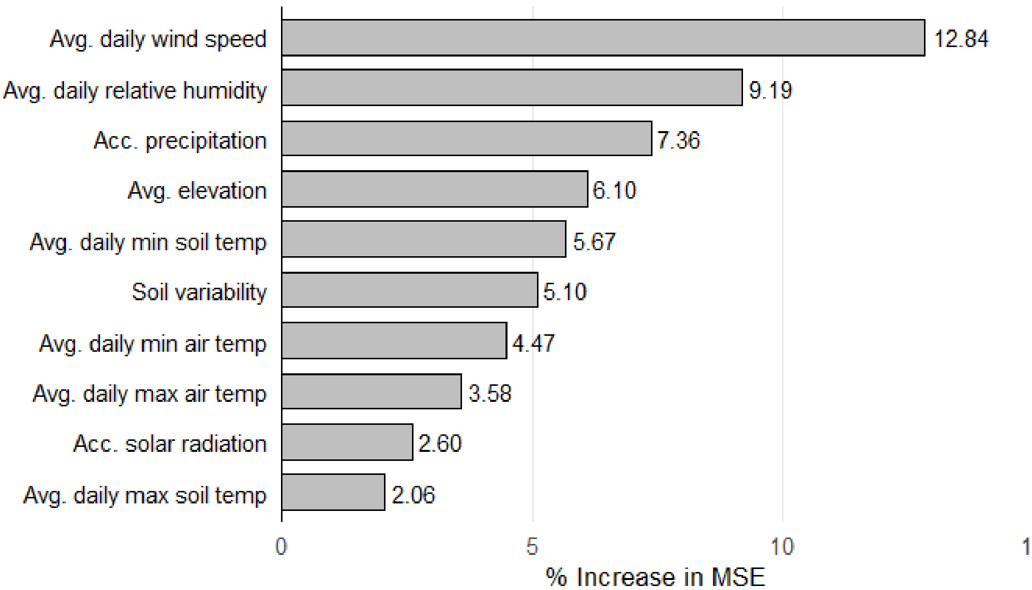


Figure 2.10 Bar chart displaying the Random Forest regression variable importance (%incMSE) for each environmental predictor of water productivity across all 32 site-years.

In summary, the results of the Pearson Correlation Matrix and Random Forest feature importance analysis concluded that across all 32 site-years, low average daily wind speed between planting and harvest was the most important environmental factor for characterizing areas where high water productivity. Additional environmental factors that had moderate significance included lower average daily soil temperatures, low average daily relative humidity, and low accumulated precipitation.

These findings support the results of the water productivity ranking (Figure 2.7). The site-years withing the Colorado Drainage Basin all had the lowest average daily wind speeds, lowest average daily soil temperatures, and lowest accumulated precipitation. Conversely, the

Northwest Kansas site-years scored the overall lowest in water productivity, and had the highest average daily wind speeds and highest average daily maximum soil temperature.

Wind can affect both yield and ET_c in maize, which are the driving factors of water productivity. On a biophysical level, the effect of wind speed on water productivity is profound. As explored by Cleugh et al. (1998), high wind speeds can damage the relatively broad leaves of the maize plant by causing abrasion, tearing, and removal of the cuticle layer from wind-driven soil particles (Cleugh et al., 1998). Damage to the maize leaves directly impacts negatively the crops ability to both photosynthesize and retain moisture. Further, high wind speeds that induce crop movement can cause growth and morphological changes such as reduced stem elongation and increased stem thickening (Cleugh et al., 1998). In severe cases, These changes caused by high wind speeds can result in an overall reduction in biomass and yield, directly affecting water productivity.

Further, in a study conducted by Davarzani et al. (2014), high wind speeds were determined to have a significant impact on soil water evaporation (Davarzani et al., 2014). Water moves from deeper, wetter layers of soil towards the drying soil surface, allowing for a continual supply of water for evaporation (Davarzani et al., 2014). When wind speeds are high, the soil surface dries faster, causing increased movement of water from wetter soils to the now dry soil surface, increasing evaporation. Additionally, high wind speeds can more quickly remove the vapor layer produced by soil temperature and water in the soil. This causes soil temperatures to rise, which subsequently allow for more water to move into the atmosphere via evaporation. These findings support the results of the Pearson Correlation Matrix and Random Forest regression feature importance, which identified soil temperature as important characteristics of high water productivity. Overall, wind speed appeared to be identified as the most important factor in water productivity in this study because of the significant impact it has on a myriad of both plant biophysical factors and soil factors that impact the physical movement of water.

2.4 Conclusions

The Chapter 1 based scaled remote sensing yield prediction approach in this study resulted in generally inaccurate yield results compared to the USDA-NASS reported data, but did manage to moderately track high and low average yield trends across all 32 site-years. Future studies may find better results by incorporating larger datasets and accounting for a more diverse range of management practices (planting dates, irrigated and non-irrigated, etc.).

Further, this study provided valuable insight into the environmental and climatic factors that most impact water productivity. Low average daily wind speed was identified as having the most positive impact on water productivity due to the impact it has on both ET_c and yield limiting factors. Future studies may find incorporate more environmental factors and site-years to build a more complete characterization of water productive areas.

The implications of this study showcase the usefulness of remotely sensed datasets in the context of maize yield and water productivity monitoring. Further research into the mitigation of the impact of wind on maize could provide helpful insights into management practices, as well as the selection of maize varieties more adapted to high wind areas.

References

Agricultural Marketing Service. (2024). *Grain Transportation Report*. https://www.ams.usda.gov/sites/default/files/media/GTR04252024.pdf

Barnes, E., Colaizzi, P., Haberland, J., & Waller, P. (2000). *Coincident detection of crop water stress, nitrogen status, and canopy density using ground based multispectral data*. https://www.researchgate.net/publication/43256762

Blaes, X., Defourny, P., Wegmüller, U., Vecchia, A. Della, Guerriero, L., & Ferrazzoli, P. (2006). C-band polarimetric indexes for maize monitoring based on a validated radiative transfer model. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(4), 791–799. https://doi.org/10.1109/TGRS.2005.860969

Boiarskii, B. (2019). Comparison of NDVI and NDRE Indices to Detect Differences in Vegetation and Chlorophyll Content. *JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES*, *spl1*(4). https://doi.org/10.26782/jmcms.spl.4/2019.11.00003

Brisco, B., Brown, R. J., Gairns, J. G., & Snider, B. (1992). *Temporal Ground-Based Scatterometer Observations of Crops In Western Canada*.

Cleugh, H. A., Miller, J. M., & Böhm, M. (1998). *Direct mechanical effects of wind on crops*.

Colorado Agriculture and Meteorological Network. (2026). *Daily Summaries*. https://coagmet.colostate.edu/

Davarzani, H., Smits, K., Tolene, R. M., & Illangasekare, T. (2014). Study of the effect of wind speed on evaporation from soil through integrated modeling of the atmospheric boundary layer and shallow subsurface. *Water Resources Research*, *50*(1), 661–680. https://doi.org/10.1002/2013WR013952

Esri. (2024). *ArcGIS Pro*. Environmental Systems Research Institute. https://www.esri.com/arcgis/products/arcgis-pro

Esri. (2025a). *Aggregate (Spatial Analyst)*. ArcGIS Pro Tool Reference. https://pro.arcgis.com/en/pro-app/3.4/tool-reference/spatial-analyst/aggregate.htm

Esri. (2025b). *Focal Statistics (Spatial Analyst)*. ArcGIS Pro Tool Reference. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/focal-statistics.htm

Esri. (2025c). *Kriging (Spatial Analyst)*. ArcGIS Pro Tool Reference. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/kriging.htm

Esri. (2025d). *Raster to Point (Conversion)*. ArcGIS Pro Tool Reference. https://pro.arcgis.com/en/pro-app/3.4/tool-reference/conversion/raster-to-point.htm

Esri. (2025e). *Reproject*. ArcGIS Pro Python. https://pro.arcgis.com/en/pro-app/latest/arcpy/image-analyst/reproject.htm

Esri. (2025f). *Sample (Spatial Analyst)*. ArcGIS Pro Tool Reference. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/sample.htm

Esri. (2026a). *Mosaic To New Raster (Data Management)*. ArcGIS Pro Tool Reference. https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/mosaic-to-new-raster.htm

Esri. (2026b). *Pairwise Clip (Analysis)*. ArcGIS Pro Tool Reference. https://pro.arcgis.com/en/pro-app/latest/tool-reference/analysis/pairwise-clip.htm

Esri. (2026c). *Raster Calculator (Spatial Analyst)*. ArcGIS Pro Tool Reference. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/raster-calculator.htm

European Space Agency. (2025). *Copernicus Browser*. https://browser.dataspace.copernicus.eu/

European Space Agency Earth Observation Portal. (2026a). *Copernicus: Sentinel-1*. Satellite Missions Catalogue. https://www.eoportal.org/satellite-missions/copernicus-sentinel-1

European Space Agency Earth Observation Portal. (2026b). *Copernicus: Sentinel-2*. Satellite Missions Catalogue. https://www.eoportal.org/satellite-missions/copernicus-sentinel-2

FAO. (2022). *The dimensions of water productivity*.

Fong, T. Y., Huang, Y. F., Chin, R. J., & Koo, C. H. (2025). Advancing evapotranspiration estimation with remote sensing and artificial intelligence – A review. In *Agricultural Water Management* (Vol. 322). Elsevier B.V. https://doi.org/10.1016/j.agwat.2025.110023

Food and Agriculture Organization of the United Nations. (2025). Agricultural production statistics. *FAOSTAT Analytical Briefs*, (No. 121). https://doi.org/10.4060/cd8035en

Food and Agriculture Organization of the United Nations. (2026). *Maize*. FAO. https://www.fao.org/land-water/databases-and-software/crop-information/maize/en/

Foreign Agriculture Service. (2024). *2024 United States Agricultural Export Yearbook*. https://www.fas.usda.gov/sites/default/files/2025-05/2024-Final.pdf

Ghosh, P., Mandal, D., Wilfling, S., Hollberg, J., Bargiel, D., & Bhattacharya, A. (2022). Synergy of optical and synthetic aperture radar data for early-stage crop yield estimation: a case study over a state of Germany. *Geocarto International*, *37*(25), 10743–10766. https://doi.org/10.1080/10106049.2022.2039306

Gitelson, A., Kaufman, Y., & Merzlyak, M. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, *58*, 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7

Google Earth Engine. (2025). *Google Earth Engine*. https://code.earthengine.google.com/

Gupta, D. K., Prashar, S., Singh, S., Srivastava, P. K., & Prasad, R. (2022). Introduction to RADAR remote sensing. In *Radar Remote Sensing: Applications and Challenges* (pp. 3–27). Elsevier. https://doi.org/10.1016/B978-0-12-823457-0.00018-5

Guttman, N. B., & Quayle, R. G. (1996). *A Historical Perspective of U.S. Climate Divisions*.

Hoeft, R. G., Aldrich, S. R., Nafziger, E. D., & Johnson, R. (2000). *Modern corn and soybean production*. https://api.semanticscholar.org/CorpusID:127398750

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1), 195–213. https://doi.org/https://doi.org/10.1016/S0034-4257(02)00096-2

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Hunt, K. A., Beeson, P., Bowman, M., Wallander, S., & Williams, R. (n.d.). *Crop Sequence Boundaries (CSB): Delineated Fields Using Remotely Sensed Crop Rotations*.

Kansas Mesonet. (2026). *Kansas Mesonet Historical Data*. mesonet.k-state.edu/weather/historical

Kansas State University Agricultural Experiment Station And Cooperative Extension Service. (2007). *Corn Production Handbook*.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22. https://CRAN.R-project.org/doc/Rnews/

Luo, P., Liao, J., & Shen, G. (2020). Combining spectral and texture features for estimating leaf area index and biomass of maize using sentinel-1/2, and landsat-8 data. *IEEE Access*, *8*, 53614–53626. https://doi.org/10.1109/ACCESS.2020.2981492

Mandal, D., Kumar, V., Ratha, D., Dey, S., Bhattacharya, A., Lopez-Sanchez, J. M., McNairn, H., & Rao, Y. S. (2020). Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data. *Remote Sensing of Environment*, *247*. https://doi.org/10.1016/j.rse.2020.111954

Marques Ramos, A. P., Prado Osco, L., Elis Garcia Furuya, D., Nunes Gonçalves, W., Cordeiro Santana, D., Pereira Ribeiro Teodoro, L., Antonio da Silva Junior, C., Fernando Capristo-Silva, G., Li, J., Henrique Rojo Baio, F., Marcato Junior, J., Eduardo Teodoro, P., & Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, *178*. https://doi.org/10.1016/j.compag.2020.105791

Melton, F. S., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J. B., Kilic, A., Senay, G. B., Volk, J., Hain, C., Johnson, L., Ruhoff, A., Blankenau, P., Bromley, M., Carrara, W., … Anderson, R. G. (2022). OpenET: Filling a Critical Data Gap in Water Management for the Western United States. *Journal of the American Water Resources Association*, *58*(6), 971–994. https://doi.org/10.1111/1752-1688.12956

Michon Scott. (2019, February 19). *National Climate Assessment: Great Plains' Ogallala Aquifer drying out*.

Nasirzadehdizaji, R., Sanli, F. B., Abdikan, S., Cakir, Z., Sekertekin, A., & Ustuner, M. (2019). Sensitivity analysis of multi-temporal Sentinel-1 SAR parameters to crop height and canopy coverage. *Applied Sciences (Switzerland)*, *9*(4). https://doi.org/10.3390/app9040655

National Agricultural Statistics Service. (n.d.). *Data Visualization by Commodity*. United States Department of Agriculture. Retrieved February 22, 2026, from https://www.nass.usda.gov/Data_Visualization/Commodity/index.php

National Centers for Environmental Information. (n.d.). *U.S. Climate Regions*. National Oceanic and Atmospheric Administration. Retrieved February 19, 2026, from https://www.ncei.noaa.gov/access/monitoring/reference-maps/us-climate-regions

National Centers for Environmental Information. (2017, July 10). *CONUS Climate Divisions*. NOAA. https://www.ncei.noaa.gov/pub/data/cirs/climdiv/

National Centers for Environmental Information. (2026). *Climate at a Glance Statewide Time Series*. NOAA. https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series

OpenET, Inc. (2026). *OpenET Ensemble Monthly Evapotranspiration v2.0*. Earth Engine Data Catalog. https://developers.google.com/earth-engine/datasets/catalog/OpenET_ENSEMBLE_CONUS_GRIDMET_MONTHLY_v2_0#description

Plastina, A., & Townsend, T. (2023). *World Spending on Agricultural Research and Development*. Center for Agricultural and Rural Development, Iowa State University. www.card.iastate.edu/ag_policy_review/article/?a=152

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Rouse, R. W. H., Haas, J. A. W., & Deering, D. W. (1974). *Monitoring Vegetation Systems in the Great Plains with ERTS*.

Sands, R., Meade, B., Seale, J. L., Robinson, S., Seeger, R., & United States. Department of Agriculture. Economic Research Service, sponsoring body. (2023). *Scenarios of global food consumption : implications for agriculture*. Economic Research Service, U.S. Department of Agriculture. https://doi.org/10113/8134356

Soil Survey Staff. (2026). *Gridded Soil Survey Geographic (gSSURGO) Database for Kansas and Colorado*. United States Department of Agriculture, Natural Resources Conservation Service. https://www.nrcs.usda.gov/resources/data-and-reports/gridded-soil-survey-geographic-gssurgo-database

Steele-Dunne, S. C., McNairn, H., Monsivais-Huertero, A., Judge, J., Liu, P. W., & Papathanassiou, K. (2017). Radar Remote Sensing of Agricultural Canopies: A Review. In *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (Vol. 10, Number 5, pp. 2249–2273). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/JSTARS.2016.2639043

Sudduth, K. A., & Drummond, S. T. (2007). Yield Editor: Software for Removing Errors from Crop Yield Maps. *Agronomy Journal*, *99*(6), 1471–1482. https://doi.org/https://doi.org/10.2134/agronj2006.0326

United Nations Development Programme (UNDP). (2026). *Algeria*. UNDP Climate Change Adaptation Portal. https://www.adaptation-undp.org/explore/arab-states/algeria

United States Geological Survey. (2026). *TNM Download (v2.0)*. GIS Data Download. https://apps.nationalmap.gov/downloader/

USDA Economic Research Service. (2025, June 18). *Crop Sequence Boundaries*. USDA-NASS. https://www.nass.usda.gov/Research_and_Science/Crop-Sequence-Boundaries/index.php

USDA Economic Research Service. (2026, January 27). *Irrigation & Water Use*. https://www.ers.usda.gov/topics/farm-practices-management/irrigation-water-use

USDA-NASS. (2023). *THE YIELD FORECASTING PROGRAM OF NASS*.

USDA-NASS. (2026). *Quick Stats*. USDA-NASS Quick Stats Tool. https://quickstats.nass.usda.gov/

Zhao, R., Ma, Y., & Wu, S. (2024). A Review of the Research Status and Prospects of Regional Crop Yield Simulations. In *Agronomy* (Vol. 14, Number 7). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/agronomy14071397