Title: Convolutional Neural Networks for Corn Yield Prediction and Optimization using Time-Series Satellite Data

Methodology

The dataset comprises time-series GeoTIFF files from NASA's assimilation databases (e.g., GLDAS for SSM, MODIS for NDVI and VGI) spanning 2000-2022. We focus on data from February to August each year to align with corn growth and harvest periods in major US regions. Features include NDVI, VGI, SSM, and other vegetation/soil indices derived from optical and SAR sources. Annual corn production data (e.g., total bushels) is sourced from USDA reports for ground truth.

Preprocessing involves:

- Extracting time-series features at county or regional levels using raster processing.
- Creating temporal lagged features (e.g., 1–4-week lags) to capture seasonal variations and harvest patterns.
- Normalizing data and handling missing values via interpolation.
- Splitting into training (2000-2020) and testing (2021-2022) sets, with 2023/2024 held for prediction.
- For SAR-derived features like SSM, we apply polarimetric decompositions as in Verma et al. (2023) to enhance feature quality.

Model Architectures

We evaluate three models for regression-based prediction:

- Convolutional Neural Network (CNN)/Deep Neural Network (DNN): A 1D CNN for time-series, with convolutional layers to extract temporal patterns, followed by dense layers. Architecture: Input (time steps × features) → Conv1D (filters=64, kernel=3) → MaxPool → Flatten → Dense (units=128) → Output. Trained with Adam optimizer and MSE loss.
- Temporal Convolutional Network (TCN): Designed for sequential data, using dilated convolutions for long-range dependencies. Layers include causal convolutions with

dilations (1,2,4,8) and residual connections. This model is particularly suited to capture seasonal variations in harvest patterns.

Training and Evaluation

- Models are trained on time-series inputs to predict 1D annualized corn production. We perform k-fold cross-validation (k=5) to assess generalizability.
- Hyperparameter tuning uses Bayesian optimization. Metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score.
- We test temporal lagged features' impact by ablating them and compare model performance. Feature importance (via RF) and attention mechanisms (in TCN) highlight variables capturing seasonal dynamics.

This methodology ensures a generalizable approach for predicting 2023/2024 corn production.

70/30 split

1. Introduction (what are the factors/theory that lead to corn yield hypothesis)

Agriculture remains the backbone of the global economy, directly influencing food security, trade stability, and industrial supply chains. Among staple crops, corn (maize) holds a particularly vital position, serving as a key input for food products, livestock feed, biofuels, and a wide range of industrial applications. In the United States, corn production underpins multi-billion-dollar sectors, from food processing to renewable energy, making accurate forecasting of its yield essential for business planning and economic resilience. Reliable corn production forecasts inform decision across the agricultural value chain: farmers can optimize resource use, agribusinesses can manage procurement and logistics, policymakers can anticipate market fluctuations, and financial institutions can assess risk exposure in commodities markets.

However, traditional methods of estimating and predicting crop yields such as field surveys, farmer reports, and statistical extrapolations are often constrained by limited spatial coverage, reporting delays, and human bias. These limitations hinder the ability of businesses and policymakers to make timely, data-driven decisions, particularly in the face of increasing climate variability and market volatility. As weather extremes and resource constraints intensify, there is a growing demand for high-frequency, spatially detailed, and scalable prediction systems that can monitor crop performance in near real time.

Advances in remote sensing and machine learning have opened new possibilities for transforming how agricultural productivity is monitored and forecasted. Satellite imagery, which is captured at consistent intervals and across multiple spectral brands, offers a rich source of information about vegetation health, soil moisture, and temperature dynamics. When combined with modern data analytics and artificial intelligence, these datasets enable the development of predictive models capable of estimating crop yields at both regional and national sales with remarkable precision.

This study focuses on leveraging satellite imagery to predict corn production using data-driven, machine learning based approaches, specifically, how time-series data can be integrated into predictive mdoels to provide early and accurate yield forecasts for business and policy applications. This research aims to demonstrate how spatially, and temporally rich data can improve forecasting accuracy, enhance decision making, and mitigate financial and operational risks across agricultural supply chains.

2. Literature Review

The convergence of advanced remote sensing technologies and sophisticated machine learning algorithms is revolutionizing earth observation, particularly in the fields of agriculture and environmental science. This review synthesizes findings from recent studies to provide a comprehensive overview of the Synthetic Aperture Radar (SAR) polarimetry, optical satellite imagery analysis, and their application in crop yield prediction, soil moisture estimation, and land cover classification. Key themes include the operational capabilities of missions like Sentinel-1, the power of polarimetric decomposition techniques are for feature extraction, the challenges of cloud contamination optical data, and the increasing sophistication of machine learning models, from Random Forest to deep learning and ensemble architectures, for integrating multi-source data. The synthesis highlights a clear trend towards leveraging high resolution, multimodal data streams to build predictive models of increasing accuracy and utility for precision agriculture resource management.

Accurate and timely monitoring of agricultural systems and environmental conditions is critical for global food security, economic stability, and sustainable resource management. Traditional methods of data collection are often labor-intensive and spatially limited. Remote sensing offers a scalable solution, providing extensive spatial and temporal coverage of the earth's surface [1, 9]. This review examines two primary remote sensing modalities, Synthetic Aperture Radar (SAR), which provides all-weather, day and night imagining capabilities, and optical imagery, which captures spectral information valuable for assessing vegetation health.

A parallel revolution is occurring in data analysis, where machine learning and deep learning models have proven exceptionally effective at extracting meaningful patterns from large, complex remote sensing datasets [8]. Here we synthesize research across several key areas, the technical capabilities and applications of SAR systems, the challenges' inherent in optical remote sensing, and the diverse application of machine learning for predicting corn yield at scales ranging from intra-field to the entire US Corn Belt.

SAR technology is a powerful tool for environmental monitoring due to its ability to penetrate clouds and operate regardless of daylight [10]. The Copernicus Sentinel-1 mission, a constellation of C-band SAR satellites, exemplifies modern operation SAR capabilities. It provides continuous, reliable data for a wide range of applications, including maritime surveillance, ice monitoring, and mapping of forests, water, and soil [10].The mission offers several imaging modes with varying resolutions and swatch widths, such as the Interferometric Wide swath (IW) and extra Wide swath (EW) modes and supports dual-polarization capacities crucial for detailed analysis [10].

Fully polarimetric SAR (PolSAR) systems, such as the L-band PALSAR, transmit and receive both horizontally (H) and vertically (V) polarized signals, capturing the full scattering matrix HH, VV, HV, VH) [11]. This rich dataset reveals information about the geometric and physical properties of targets that is invisible in single polarization imagery [11] creating RGB colour composites from different polarization channels, distinct scattering characteristics of various land cover types can be visualized effectively [11].

To quantitatively extract this information, polarimetric decomposition techniques are employed to break down the complex scattering signal into fundamental components related to physical mechanisms [3]. These methods are broadly categorized and coherent and non-coherent.

Coherent decomposition assumes a single dominant scattering mechanism per resolution cell [3].The Pauli decomposition, for instance, expresses the scattering matrix as a sum of bases representing single bounce, double bounce, and volume scattering, which is useful for visualization but can lack discrimination [3].

Non coherent decomposition is a method that use second order statistics to analyse areas where multiple scattering mechanisms coexist [3]. Model based approaches like the Freeman-Durden, and Yamaguchi decompositions separate the signal into surface, double-bounce, and volume scattering components [3]. The Yamaguchi model further adds a helix scattering component to better characterize complex urban or man-made targets [3]. Eigenvector based methods like the Cloude-Pottier H/A/ α decomposition derive parameters like entropy (H) and alpha angle (α)to quantify scattering randomness and identify the dominant scattering type [3].

Hybrid models are newer approaches, like the Double Scatter Model combine deterministic and non-coherent features, interpreting each cell as a combination of the two most dominant scattering mechanisms, showing high accuracy in land cover classification [3, 9].

A key application of SAR is the estimation of surface soil moisture (SM). However, the presence of vegetation introduces volume scattering that contaminates the soil backscatter signal, compromising the accuracy of semi-empirical retrieval algorithms like the oh model [7]. A recent study introduced a novel hybrid approach to address this challenge using Sentinel-1 dual-polarimetric data. By integrating model-based polarimetric decomposition to isolate and reduce the vegetation volume contribution before applying the Oh model, the method significantly improves SM estimation accuracy [7]. This technique achieved a 25%-52% lower Root Mean Squared Error (RMSE) compared to the Oh model and, crucially, does not require in-situ plant descriptors, making it scalable for regional mapping [7].

While SAR provides structural information, optical sensors capture reflected solar radiation across multiple spectral bands, which is essential for applicants like land cover classification, crop identification and laser monitoring [9]. However, a major impediment to the usability of optical satellite imagery is contamination by clouds and their shadows, which can obscure the Earth’s surface and compromise data integrity [9] Cloud detection, or masking, is a critical preprocessing step to identify and exclude cloud-affected pixels [9] range from manual interpretation to automated algorithms [9]. Automated techniques include traditional threshold-based methods like F-mask and Sen2Cor, which apply rules to spectral bands, and more advanced machine learning and deep learning approaches that learn to identify cloud features from large, labelled datasets [9]. The development of robust, automated cloud masking is essential for creating analysis ready data at scale.

Predicting crop yield is a complex challenge influenced by genetics, environment, and management practices [8]. Machine learning has emerged as a powerful tool, capable of modelling the intricate, non-linear relationships between various predictor variables and yield [8]. Research in this area leverages diverse data sources and a wide array of ML machine learning models.

Many studies integrate multi-source satellite data for yield prediction. Successful models combine optical data from Sentine-2, thermal data (Land Surface Temperature, LST) form Landsat, and vegetation indices like EVI and LAI from MODIS, along with agroclimatic data [1, 2]. The combination of multiple indices consistently improves performance, with LST being particularly valuable for detecting early-season drought stress that optical sensors may miss [1, 2]. For monitoring within-field variability, Sentinel-2 imagery has proven effective, with the Green Normalized Difference Vegetation Index (GNDVI) showing the strongest correlation with corn yield during the R4-R6 physiological maturity stages [4].

For higher resolution analysis, Unmanned Aerial Vehicles (UAVs) provide data at the intra-field level, a scale difficult to achieve with satellites [5] study demonstrated that multitemporal UAV RGB imagery, when fed into spatio-temporal deep learning models, can predict yield with high accuracy (5.51% MAPE) [5]. Importantly, accurate predictions were still possible using shortened time-series data from early in the growing season, providing actionable insights for farmers [5] .

A variety of machine learning algorithms have been successfully applied to yield prediction, with a clear trend towards more complex and powerful models.

The Random Forest (RF) algorithm is a robust and frequently used model. It has proven effective in predicting corn yield from multi-source satellite data and was identified as the most accurate technique for monitoring within-field variability [1, 4]. One study focused specifically on optimizing RF hyperparameters (n_estimator, max_depth, etc.), demonstrating that careful tuning can significantly minimize prediction error and avoid overfitting, achieving a MAPE of just 9.04% [6].

Deep learning (DL) architectures are particularly well-suited for handling the spatial and temporal nature of remote sensing data. Convolutional Neural Networks (CNNs) excel at extracting spatial features, while Long Short-Term Memory (LSTM) networks, a type of RNN< are designed to model temporal sequences [2, 8] . The combination of these architectures is common, and LSTMs have consistently achieved top performance when predicting yield from time-series satellite data [2]. For high resolution UAV time-series data, a 3D-CNN, which applies non convolutions in both spatial and temporal dimensions, was found to outperform other spatio-temporal models like CNN-LSTM and ConvLSTM [5] .

To further boost accuracy and robustness, ensemble models combine predictions from multiple base learners [8]. A study developing ensemble CNN-DNN models found that homogenous ensembles, created by training the same baes model on different bootstrapped data samples (bagging), provided the most accurate county-level corn yield predictions [8]. These ensembles outperformed individual models and literature benchmarks by 10-16%, demonstrating the power of reducing prediction variance through model aggregation [8].

The body of research reviewed illustrates a dynamic and rapidly advancing field. SAR and optical remote sensing provide complementary information about the Earth’s surface, with SAR offering all-weather structural data and optical sensors providing rich spectral detail. The synthesis reveals a clear progression in analytical techniques, moving from simple vegetation indices to the integration of multi-source data streams from satellites, UAVs, and climate models. Machine learning is the engine driving this progress, with models evolving from established algorithm like Random Forest to highly specialized deep learning and ensemble architectures tailored for spatio-temporal data. Together, these technologies are enabling unprecedented capabilities in monitoring agricultural productivity and environmental change, providing critical tools to support precision agriculture, policy-making, and global food security.

3. Data and Materials

The bulk success of this research expedition relied heavily on quality data, not just quality data that contained the right features required for training the model, but a lot of it. For the model to learn and effectively make predictions, it needs a large amount of recorded crop yield production that spanned decades without such; temporal learning would be virtually impossible. Thankfully, USDA Agricultural Statistics Service (NASS) has done an incredible job of recording maize production across the county which they have made open source to foster more research on new innovative approaches to use the data and improve crop yield production prediction. In addition to crop yield production records, geospatial data was needed as well. Geospatial data contains so much valuable information about the land surface conditions which are necessary for the model to build a robust understanding of features such as the geological factors that have a direct or indirect influence on crop yield. A simple example is land surface temperature which provides information on the temperature of the soil surface through the span of the crop life. There is a plethora of other features which the model would need to take into consideration to build a comprehensive correlation matrix. NASA's Global Land Data Assimilation System (GLDAS) is an invaluable repository of such information which we utilized for this research project.

Customarily, data from the GLDAS comes in a different file format which is called geographical tagged image file format, more commonly known as GeoTIFF. Before this data can be used, it must be converted, and this can be achieved using free python libraries such as geopandas, rasterio, fiona, shapely, earthengine-api and a couple of others which will be listed in a requirements.txt file that'll be attached in an effort to improve accessibility and reproducibility via peer-to-peer review of this research. Due to the nature and size of the data, some configuration had to be done, one such approach was grouping the annual crop yield production from 2000-2024 according to county and there are other approaches that can be tried which is entirely at your discretion. Both datasets were combined on a composite key of country name and year where each row represents a unique county name, year, and month. With the data combined, it could now be filtered and sorted according to country and year. The now combined dataset contained features which can be categorized according to identifiers (fips, country_name, year, month), temperature (AvgSurfT_inst, Tair_f_inst, SoilTMPO_10cm_inst, ... ,Tveg_tavg), water & moisture (Rainf_tavg, Evap_tavg, SoilMoi0_10cm_insti, ..., RootMoist_inst, SWE_inst), heat & energy (Qh_tavg, Qle_tavg, Qg_tavg). For data cleaning and preprocessing, counties with reported production under

1,000 bushels each year were considered non-producing or having insignificant activity and as such, were excluded entirely from the dataset.

- **Hydro-Meteorological Data (GLDAS)**: Derived from NASA's Global Land Data Assimilation System (GLDAS v2.1), this includes 19 variables such as evapotranspiration from soil (ESoil_tavg), layered soil moisture (SoilMoi0_10cm_inst, SoilMoi10_40cm_inst, SoilMoi40_100cm_inst, SoilMoi100_200cm_inst), soil temperature profiles (SoilTMP0_10cm_inst to SoilTMP100_200cm_inst), downward longwave radiation (LWdown_f_tavg), air temperature (Tair_f_inst), wind speed (Wind_f_inst), average surface temperature (AvgSurfT_inst), total evapotranspiration (Evap_tavg), canopy interception (CanopInt_inst), sensible heat flux (Qh_tavg), ground heat flux (Qg_tavg), canopy evapotranspiration (ECanop_tavg), surface pressure (Psurf_f_inst), snowmelt (Qsm_acc), snow water equivalent (SWE_inst), surface runoff (Qs_acc), snow depth (SnowDepth_inst), and albedo (Albedo_inst). These were extracted via Google Earth Engine (GEE) at 0.25° resolution, aggregated monthly per county using zonal statistics (mean values), and filtered for the growing season. GLDAS data provide critical insights into water and energy balances, which are pivotal for modeling drought stress and vegetative growth.
- **Weather Data (PRISM)**: Sourced from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) dataset via the Oregon State University API. Features include monthly precipitation (mm), minimum/maximum/mean temperature (°C), and vapor pressure deficit (VPD) proxy (calculated as the difference between saturation and actual vapor pressure). Data were quality-controlled (QC'ed) for outliers (>3σ from county means) and unit-normalized (z-score scaling) to facilitate model convergence. Windowed aggregates (e.g., 30-day rolling means for temperature and radiation) were computed to capture intra-seasonal variability.
- **Drought Indicators (USDM)**: From the U.S. Drought Monitor (USDM), monthly percentages of county area and days under D2+ severity (intense drought or worse) were included. These were aggregated from weekly raster data using ArcGIS zonal tools, providing a proxy for prolonged water stress impacting reproductive stages.
- **Soil Properties (gSSURGO)**: Static features from the Gridded Soil Survey Geographic Database (gSSURGO) include available water capacity (AWC, cm/cm), organic matter percentage (OM%), and a drainage class score (1–7 scale, where 1=excessively drained, 7=very poorly drained). These were averaged per county using USDA's Soil Data Development Tools, reflecting inherent soil productivity and water-holding potential.
- **Cropping Patterns (CDL)**: Annual corn fraction (proportion of county area planted to corn) from the USDA Cropland Data Layer (CDL), derived via GEE at 30m resolution. This captures land-use intensity and potential competition for resources.
- **Markets and Logistics**: Economic drivers include December corn futures prices (January–April average and volatility, from CME Group API), Minnesota pre-planting cash prices (prior-harvest averages), state basis proxy (difference between cash and

futures prices), and geodesic distance to the nearest ethanol plant (km, calculated using Haversine formula from plant coordinates in USDA's Biofuels Atlas). These reflect market incentives and logistical constraints influencing planting decisions.

- **Local Economy (DEED County Profiles)**: Socio-economic variables from the Minnesota Department of Employment and Economic Development (DEED), including population, median age, unemployment rate (%), cost-of-living hourly wage, regional median wage, and median wages for selected industries (e.g., agriculture, manufacturing). Annual profiles were scraped via web tools and interpolated for missing years using linear methods.

Ground-truth corn yields (bushels per county-year) were sourced from USDA National Agricultural Statistics Service (NASS) Quick Stats API, log-transformed to handle skewness and stabilize variance. The final dataset yielded 1,718 samples (87 counties × ~20 years, with minor gaps imputed via k-nearest neighbors). Features totaled 44 after engineering interactions (e.g., evap_moist_interact = Evap_tavg × RootMoist_inst) and trends (e.g., year_trend for technological progress).

5. Results

The model demonstrated strong predicted performance, achieving an R^2 value of 0.89 using log-transformed yield data and 0.90 after scaling adjustments. The minor difference between the log and scaled transformations reflects the rounding and normalization effects rather than substantive variation in model performance. Overall, these results indicate a high correlation between satellite derived features and observed corn prediction, underscoring the potential of using machine learning for yield estimation.

The remaining 10% gap in accuracy can largely be attributed to external and region-specific factors that are not fully captured by satellite imagery alone. For instance, each county in Minnesota exhibits unique agricultural and environmental patterns influenced by local soil quality, irrigation practices, and microclimates. Moreover, certain counties report negligible or zero corn production due to land-use differences, forest coverage, and non-agricultural economic activities. These variations introduce noise that limits the ability to perfectly generalize across all regions.

Spatial analysis of the results reveals that there is a clear latitudinal gradient in predictive accuracy. Counties in southern Minnesota showed significantly higher prediction of accuracy than those counties in the north. This pattern reflects underlying agricultural reality, where southern counties account for a larger share of the state's corn production, benefiting longer growing seasons, higher corn fertility, and more favorable climatic conditions. In contrast, northern Minnesotan counties are characterized by shorter growing periods, lower average temperatures, and greater forest coverage, all of which constrain large-scale corn cultivation.

Additionally, proximity to processing facilities and transportation networks plays a role in the spatial distribution of corn production. Southern counties tend to be more densely located near supply chain infrastructure, which supports higher production intensity and consistent reporting. These factors, while indirectly captured by satellite data through vegetation patterns, contribute to the model's stronger performance in southern regions.

While our model achieved robust predictive accuracy, its performance variation across highlights the importance of integrating contextual and economic variables such as land use data, supply chain proximity, and climate indices to capture the remaining sources of variance and further improve model generalizability.

6. Discussion

Corn production in Minnesota varies considerably across counties, reflecting the combined influence of diverse economic, logistical, and environmental factors. While satellite derived features such as vegetation indices and temperature patterns provide valuable insights into crop health and potential yield, they cannot account for the localized socio economic and infrastructural differences that shape production vs. Outcomes. Each county's capacity to produce corn is influenced by multiple interrelated variables, including the local economy, which determines farmer's ability to invest inputs and technologies, and market demand, which dictates the extern tot which corn cultivation is prioritized. Equally important are the country's connections to the 5 key supply chains that sustain corn movement: storage facilities, feed mills (livestock feed), ethanol plants (fuel), processing centers (for direct human consumption), and export terminal (exporting to other states and/or countries). Limited access to efficient supply chains may discourage large-scale cultivation, as farmers without reliable distribution channels are less likely to expand production.

This complexity makes it challenging to predict corn yields solely from satellite and climatic data, as these datasets primarily reflect biophysical conditions rather than economic or infrastructural constraints. Interestingly, the analysis revealed that county level attributes exhibit the highest correlation with corn production, surpassing the predictive strength of individual environmental variables. This suggests that regional context, capturing both human and structural dimensions, plays a critical role in determining agricultural output.

Furthermore, biological and ecological dynamics, such as the migration of pests and use of pesticides, also contribute to yield variability. These factors fluctuate spatially and temporally, influencing crop health in ways that may not be directly visible in satellite imagery. For example, increased pest pressure in certain regions may reduce yields despite favorable weather and vegetation conditions, while the intensity of pesticide application may mitigate losses but alter crop vigor and spectral characteristics. Incorporating these external influences into future models could enhance prediction of accuracy and provide a more holistic understanding of the drivers of corn production across different regions.

7. Conclusion

One key takeaway from this research is that having access to farmer planting data could significantly improve the ability to predict harvest outcomes. This is because harvest predictions depend primarily on environmental and weather conditions, whereas production predictions are influenced by additional factors such as management practices and resource inputs.

Minnesota's weather patterns are highly localized, which presents challenges in generalizing these findings to other states. However, with continued research, the predictive accruacy could be enhacned, particularly by refining models using more localized environmental data.

Future research is moving in the right direction toward developing a reliable method to predict harvest outcomes in Minnesota using satellite data. While the methodology has the potential to be generalized to other states, the results of this study are not directly generalizable due to the unique regional characteristics.

Going forward, obtaining detailed planting information, such as the number of seeds or corn stalks planted and the specific planting dates would be invaluable. These data could further strengthen the predictive models by providing more context on crop growth conditions from the very beginning of the planting cycle.

References

[1] Desloires, J. et al. 2024. Early Season Forecasting of Corn Yield at Field Level from Multi-Source Satellite Time Series Data. *Remote Sensing*. 16, 9 (May 2024). https://doi.org/10.3390/rs16091573.

[2] Ji, Z. et al. 2022. Prediction of Corn Yield in the USA Corn Belt Using Satellite Data and Machine Learning: From an Evapotranspiration Perspective. *Agriculture (Switzerland)*. 12, 8 (Aug. 2022). https://doi.org/10.3390/agriculture12081263.

[3] Karachristos, K. et al. 2024. A Review on PolSAR Decompositions for Feature Extraction. *Journal of Imaging*. Multidisciplinary Digital Publishing Institute (MDPI).

[4] Kayad, A. et al. 2019. Monitoring within-field variability of corn yield using sentinel-2 and machine learning techniques. *Remote Sensing*. 11, 23 (Dec. 2019). https://doi.org/10.3390/rs11232873.

[5] Nevavuori, P. et al. 2020. Crop yield prediction using multitemporal UAV data and spatio-temporal deep learning models. *Remote Sensing*. 12, 23 (Dec. 2020), 1–18. https://doi.org/10.3390/rs12234000.

[6] Prasetyo, T.A. et al. 2024. Parameter Optimization of the Random Forest Algorithm for Predicting Corn Yields in Toba Regency. *7th International Seminar on Research of Information Technology and Intelligent Systems: Advanced Intelligent Systems in Contemporary Society, ISRITI 2024 - Proceedings* (2024), 1025–1029.

[7] Roy, P.D. et al. 2025. Retrieval of Surface Soil Moisture at Field Scale Using Sentinel-1 SAR Data. *Sensors*. 25, 10 (May 2025). https://doi.org/10.3390/s25103065.

[8] Shahhosseini, M. et al. 2021. *Corn Yield Prediction with Ensemble CNN-DNN*.

[9] Singh, R. et al. 2025. Cloud Detection Methods for Optical Satellite Imagery: A Comprehensive Review. *Geomatics*. 5, 3 (Jun. 2025), 27. https://doi.org/10.3390/geomatics5030027.

[10] 2025. *Copernicus: Sentinel-1*.

[11] 1997. *Polarimetric observation by PALSAR*.

Methodology

Purpose: Explain how the research was conducted and why specific methods were chosen.

Key Questions:

- What methods, models, or analytical approaches were used?
- Why were these methods appropriate for your research question?
- How were variables measured or defined?
- What assumptions underlie the approach?
- How was validity or reliability ensured?
- How did we decide on what models?
- Data cleaning process

Feature selection:

- County name (one hot encoded)
- Root moisture
- Evaporation moisture
- Tveg vegetation
- Year to year trend (corn production) for model analysis
- Lwdown
- Tair temperature of air
- Soil moisture

Using PCA analysis correlation dimension anlaysis (insert graphs, etc)