Harvesting Trends Using Time Series Insights

Dallin Anderson, Joseph DuToit, Oscar Escobar, Abraham Harris, Joseph Humpherys

Abstract. Increased variability and severity of weather patterns due to climate change poses a significant challenge to agriculture in developing countries. We seek to understand the influence of weather on agricultural production and food security in developing countries using state space models, smoothing, and regression techniques. In particular, we predict Normalized Difference Vegetation Index (NDVI) and food price indices. We demonstrate the ability to predict seasonal weather patterns and vegetation levels in Ukraine, Chad, and the Philippines. We find a strong correlation between weather and NDVI and can make accurate predictions of NDVI from regional weather. Our methods for forecasting food price indices, on the other hand, prove less successful.

1. Problem Statement and Motivation

The year 2024 was the hottest on record surpassing the previous holder, 2023, by 0.18 degrees Fahrenheit, and boosting Earth's average temperature to 2.63 degrees Fahrenheit above pre-industrial levels [NOA25]. This global warming has triggered more erratic and dramatic natural disasters such as floods, droughts, and wildfires with especially devastating consequences on the developing world. Without the resources, infrastructure, and influence to attenuate the effects of climate change, these nations are often literally left out to dry. Researchers estimate that up to 85 percent of the global GDP loss due to climate change will be borne by developing regions [Ado24].

One reason for the disproportionate effects of climate change on developing nations is their dependence on agricultural production, both for export and subsistence. Unpredictable, extreme weather patterns complicate farming techniques and put crop yields in jeopardy. In this project we attempt to capture the influence of weather on agricultural production and food security in developing nations. We employ data-driven methods to model and forecast weather patterns and then predict their influence on agricultural metrics. Specifically, we seek to answer the questions:

(1) Can we accurately predict weather patterns?
(2) Can we model how weather patterns affect local food price and vegetation growth?
(3) How might predicted changes in climate impact agricultural production?

Date: April 12, 2025.

A successful model has great potential to prepare farmers for potential climate hazards and inform preventative measures for limiting food instability in developing countries.

Our work adds to a rich conversation of current research on this relevant problem. Several other publications take a similar approach using NDVI to measure agricultural yields [Geo] [GIS25] [oA25]. In a study related to ours, Brown et al. use NDVI and economic factors to predict local food pricing with a state space model showing that NDVI anomalies have a large effect on food prices in certain areas of the world [BK15]. Furthermore, [Unk17] use weather feature in addition to NDVI measurements to predict wheat yield in Pakistan. Others studies extend this work implementing machine learning techniques for similar purposes in other regions [AKA$^+$24]. Our focus differs from these studies in our generalizable approach which can be applied to any country with available weather, price, and economic data. In this work we applied our techniques to the Philippines, Ukraine, and Chad to explore environments with contrasting climates.

2. Data

Our project required data from several trustworthy sources to give us insights on weather, food prices, and vegetation. We obtained weather data from the National Oceanic and Atmospheric Administration (NOAA) spanning from January 1, 2001 to December 31, 2024. Each of our case studies used data from different weather stations. For the Philippines we used Ilocos Sur, for Ukraine we used Kharkiv, and for Chad we used Am Timan. Stations were selected based on data completeness and proximity to agricultural hubs. The data is comprised of measurements of daily maximum temperature (Fahrenheit), average temperature (Fahrenheit), minimum temperature (Fahrenheit), precipitation (inches), maximum wind speed (miles per hour), and dew point (Fahrenheit).

The weather data was mostly complete, with the one exception of Ukraine's weather data from early 2022 coinciding with the Russian conflict. We consequently only used Ukrainian data from before 2022. For the few other values that were missing we imputed using the value of the previous day and forward filling. Our data has high enough resolution that we lose little information in forward filling, and it is reasonable to suspect that weather does not drastically differ from day to day. One significant limitation of our approach is that we analyze only local data and do not consider data in the region as a whole. Irrigation farming often exhibits complex dependencies on weather patterns in distant regions from which water is sourced rather than local weather only. Unfortunately, examining these fascinating relationships is outside the scope of this project.

The food price data was sourced from the World Bank's Microdata library that estimates monthly prices in local markets using actual measurements and predictive modeling from economic features. Price ratio estimates are given from Jan 2007 up to the present. We specifically used the study's food price ratios for a general representation of food price behavior. We used data from the Ilocos Sur market in the Philippines and from the Am Timan market in Chad. Again, we selected markets based on distance from weather stations and agricultural hubs. We multiplied the food price ratios by 100 to convert them to food price indices. Note that food price index data was not available for Ukraine, and we only studied Ukrainian weather and vegetation data, as we motivate later.

Finally, we procured NDVI data from the World Food Programme, who computed the data from NASA's Moderate Resolution Imaging Spetroradiometer collection. The data begins on July 7, 2002 and ends March 3, 2025 in 10 day increments. NDVI is locally aggregated to give regional estimates of plant health. In every case we selected the region immediate to the corresponding weather station.

The first step to temporally align our data was to match the start and end dates for any comparison. Next, since our data was sampled at different rates, we used monthly average aggregation to match data. In every case we train on the first 80 percent of dates and test on the last 20 percent chronologically, with the exception our NDVI predictions in which we used a 70-30 split. To avoid data leakage, we only fit models and estimate hyperparameters on our training set.

3. Methods

Our preliminary goal was to develop an accurate model for predicting all of our weather features. We used the classical decomposition of a time-series $X(t)$ into trend $T(t)$, seasonal $S(t)$, and covariance stationary residuals $R(t)$ given by $X(t) = T(t) + S(t) + R(t)$. We first extracted the linear trend from the data. Then we isolated the seasonal component by assuming an annual period and computing the daily average across the years. Lastly, we experimented with training ARMA and SARIMAX (seasonally adjusted) models on the residuals of each feature. The residuals were not perfectly covariance stationary, and we wanted to see if SARIMAX could capture higher variance in different seasons. To get the best performance we did a grid search on parameters p and q and selected the model with the lowest AIC. At prediction time, we add the trend, seasonal component, and residual prediction to reconstruct an estimate of the time series.

After successfully implementing weather prediction, we attempted to estimate local food prices in Ilocos Sur. We first applied exponential smoothing as we theorized that the most recent food price index would impact future observations independent of weather patterns. We employed simple exponential smoothing, Holt smoothing, and Holt-Winters smoothing. These three techniques stem from the equation $\hat{y}_t = \alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}$ where α

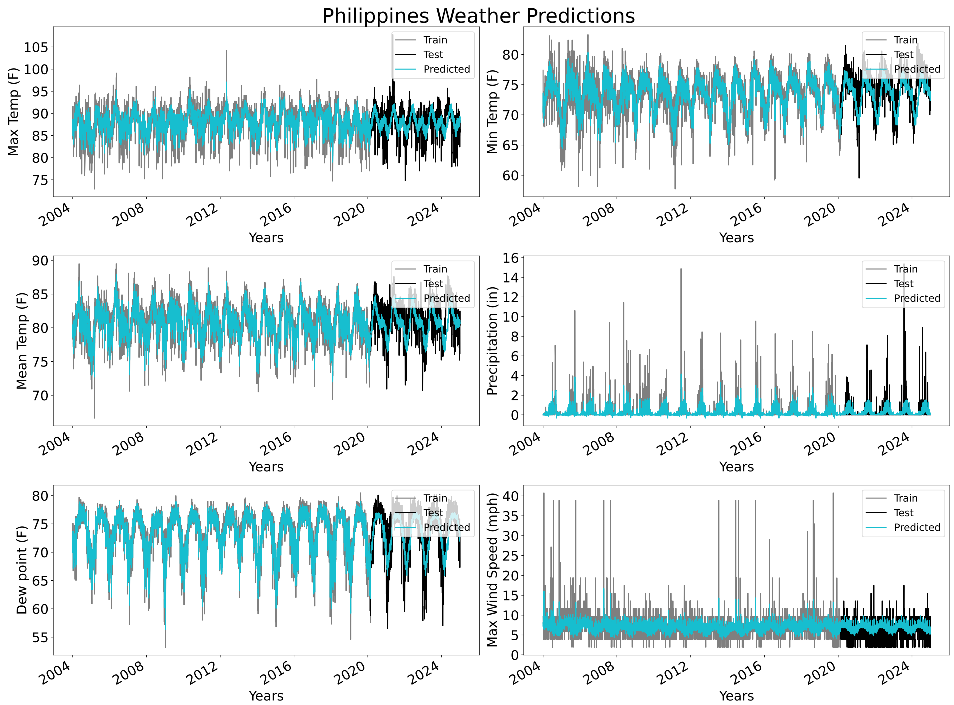

FIGURE 1. Philippines weather data and predictions from January 1, 2001 to December, 2024. The training data is colored black, testing data is colored gray, and the predictions are colored cyan. Our weather model accurately captures and forecasts weather trends and can be easily adapted to new data.

controls how much the previous estimate of the true observation y_{t-1} controls our future observation. We then add terms to incorporate trends (Holt) and trend and seasonality (Holt-Winters).

We then introduced weather data into our modeling efforts. We first plotted each weather trend against food price data and examined the correlation. We found that local food prices did not exhibit a strong connection with local weather data. We discuss this finding further in our analysis. Even though predicting prices did not seem feasible with our data, we hypothesized that we might still be able to understand the agricultural impact of weather patterns by observing plant growth via NDVI. Plant growth is more directly linked to weather and seasonal patterns and less susceptible to outside, unmodeled economic forces.

To this end, we plotted NDVI against weather patterns to understand their relation. First, we tried comparing the residuals of our model against NDVI data, but we did not see significant correlation. We next tried comparing plant growth in Ukraine and Chad – countries with seasonal cycles that strongly influence growing seasons. As expected, we recognized seasonal correlations in Ukraine and Chad, but not in the Philippines, where weather has less seasonal variation and plants grow perennially. This suggests that our method might be best restricted to regions where NDVI and growing seasons can be directly linked to agricultural activity. Although our project originally sought to understand the influence of extreme weather on agriculture, our data was not sufficient to model this phenomenon. Instead, we decided to focus on the influence of seasonal trends on plant growth in countries with intelligible seasonal trends.

We proceeded by constructing and fitting a nonlinear least-squares model of the form

$$f(\mathbf{X}, \mathbf{\Theta}) = \sum_{i=1}^{n} \theta_i^{scale} g_i(\mathbf{X}_i, t - \theta_i^{shift}) + \theta^{bias}$$

where $\mathbf{X}$ is a data matrix with n weather features over a time interval as columns (denoted $\mathbf{X}_i$), $\mathbf{\Theta}$ is the set of parameters

$$\mathbf{\Theta} = \{(\theta_i^{shift}, \theta_i^{scale}) : i \text{ in } 0, 1, \ldots n\} \bigcup \{\theta^{bias}\},$$

and $g_i(\mathbf{X}_i, t)$ is the an interpolation of the feature $\mathbf{X}_i$ over the specified time interval. The interpolations g_i were computed using `scipy`'s `CubicSpline` function. The model was fit on the training data using `scipy`'s `minimize` function with L-BFGS as the optimization algorithm. The predictions of the model can be seen in 6. The intuition behind this model was that seasonal patterns such as rainfall and temperature may have yearly periods, but be out of phase with growing patterns, even though these weather patterns strongly influence plant growth. For example, a year with heavy snowfall would not be directly correlated with plant growth since plants do not grow in cold winter temperatures. However, the increase in available moisture would encourage significant growth later on. Our model attempts to identify these relations.

Unfortunately, the design of this model is admittedly flawed and given more time and data analysis, it would be important to correct. In particular, the model assumes a linear relation between phase-corrected weather features and plant growth, which is certainly not the case. Extremes in weather conditions would inhibit plant growth, which is not expressed in the model. Initially we tried fitting a model with Gaussian shape to penalize extreme temperatures, but the model performed very poorly. In the end, we decided to use the linear model for subsequent simulations, but the design of this model warrants more thought and experimentation.

Finally, we attempted to simulate the effects of climate change on plant growth by transforming weather features to match monthly climate change predictions published by the World Bank for the year 2060 [Wor25]. This amounts to shifting the seasonal trend to match the predicted change in temperature and applying noise in accordance with the forecasted distribution for each country. By sampling 100 weather trajectories, we estimate probable future weather patterns. We then use our nonlinear least-squares model to predict NDVI. Although the aforementioned limitations of our model largely render our predictions untenable, we hope that using this technique with more realistic NDVI models could provide a clearer picture of future challenges to agriculture in developing countries.

4. Results and Analysis

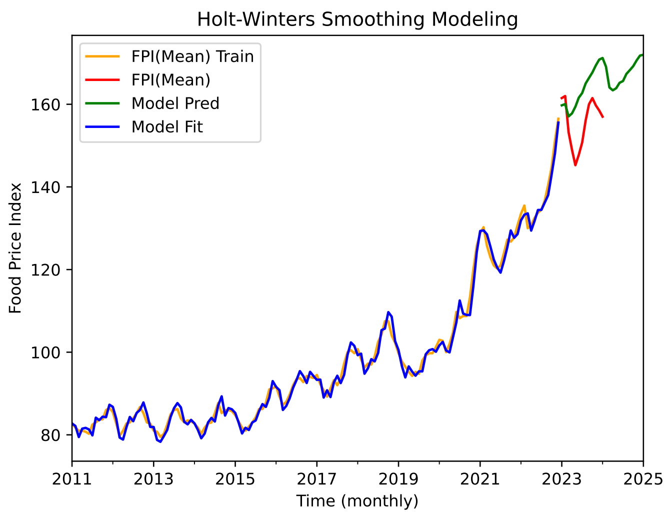

Figure 2. Predicting food price index in the Philippines for the next 2 years.

In answer to our first question, we found little correlation between weather patterns and local food prices in the Philippines and Chad. We suspect that greater economic forces outweigh the influence of weather, especially since our data was in part estimated using economic trends. It may take a more nuanced approach involving features that we did not consider to uncover if a connection between the two actually exists.

Even with Holt-Winters smoothing, which accommodates for trends and seasonality, we could not accurately predict the food prices. Our smoothing predictions follow both the upward trend in food price index as well periodic rises and drops (see Fig. 2). Nonetheless, the values deviated from the true data.

Although our results show the strengths of our approach for short term weather modeling, they also demonstrate where our methods could be refined and highlight the need for further research.

As mentioned in the methods section, we found little relationship between the residuals of our weather predictions and plant growth. This suggests that identifying the effects of increasingly frequent severe weather events is

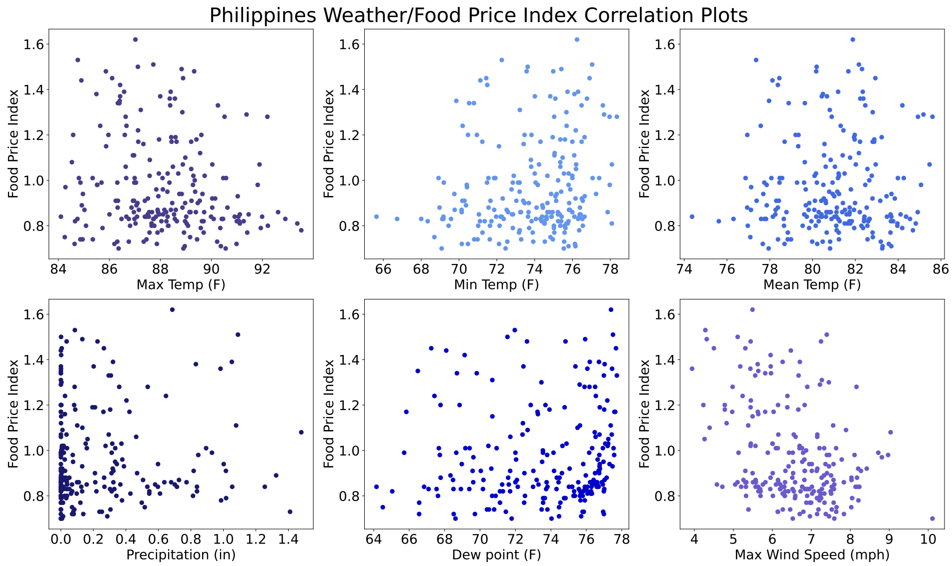

Figure 3. Correlation plots between weather features. Weather and food price show little correlation.

a difficult problem and requires a different approach. A future study may be interested in first isolating anomalies in weather and then carefully exploring changes in vegetation afterward on a case-by-case basis. However, we did identify the expected seasonal influence of weather on plant growth in Ukraine and Chad. In Ukraine, we see the obvious seasonal correlation between temperature and growth (plants do not grow in cold winter conditions) 4. However in Chad we see the interesting out of phase effects of temperature on growth (plants grow after it has been hot and the rains begin) 5.

Next, we demonstrate the performance of our nonlinear least-squares model as seen in 6. Despite the different seasonal trends in Ukraine and Chad, our model is able to accurately predict NDVI on the test set in both cases. Interestingly, around 55 months, our model correctly predicts a season with less decrease in NDVI suggesting that although our model is flawed, it may still be useful.

Finally, we generate probable future random weather data using our weather predictions and the predicted changes in climate sourced from the World Bank. Interestingly, the sampled NDVI pattern suggests slightly longer growing periods which corresponds with warmer weather in the region. When we performed the same analysis for Chad, we noticed a great dependence of the growing season on precipitation.

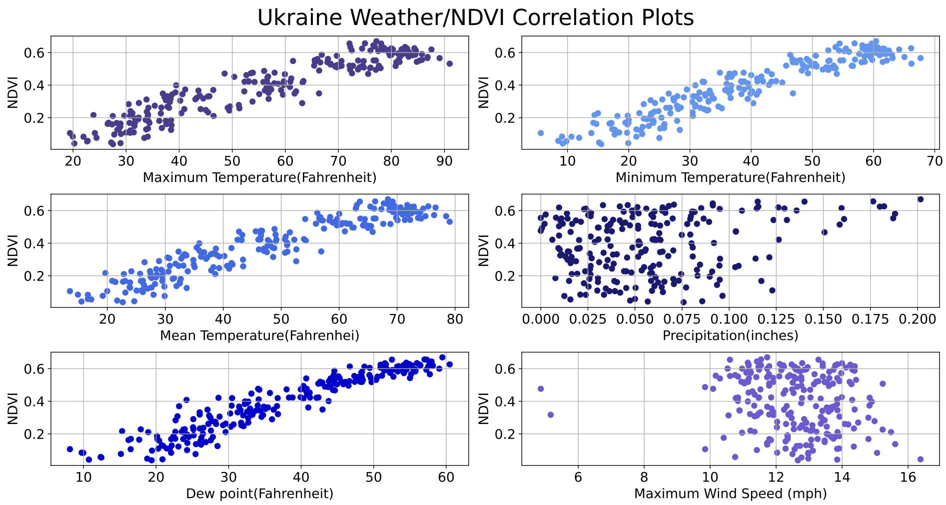

FIGURE 4. Correlation plots between Ukraine weather data and NDVI. Note the strong influence of seasonal temperature on plant growth.

5. Ethical Implications

With any project that involves gathering data and building models, it is important to consider the ethics involved. All of the data we have gathered is publicly available, contains no personal information, and describes large-scale phenomenons, so we are not concerned about any privacy or other issues regarding data collection.

However, there are several ethical considerations to take into account with how this model could be deployed and used. For example, one could use this model to predict agricultural production. Thus, if the model predicts a decrease in greenery at some point in the growing season, this could indicate potential food shortages, which could potentially lead to higher food prices. This could lead food importers, exporters, suppliers, or sellers with access to this model to proactively change prices or change their deliveries or orders, thus creating a somewhat self-fulfilling cycle of increased food price as a result of the model.

Another similar problem could result if governments or charities used models like this to determine where to send humanitarian aid. If weather is such that a certain area seems likely to experience agricultural loss, a wealthier nation may decide to send aid such as food or resources. This could have unforeseen consequences such as decreased demand for local food, potentially negatively impacting local farmers or suppliers. Thus while we hope

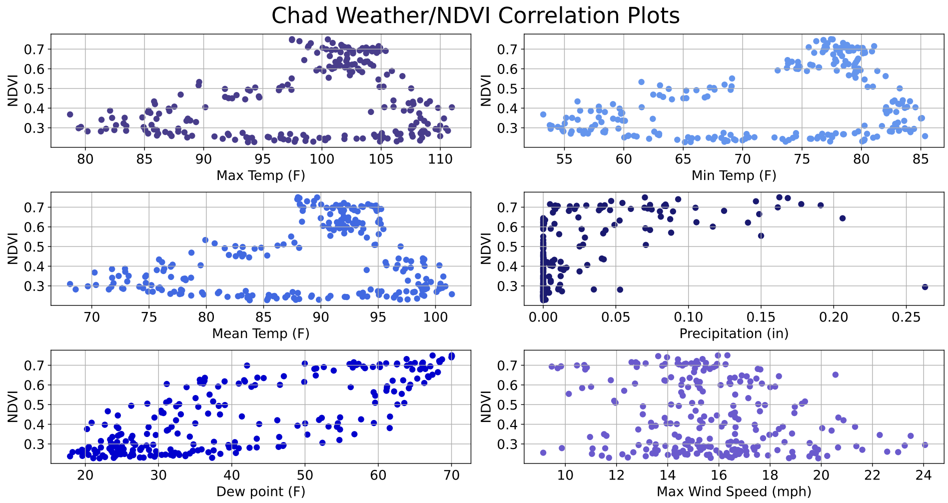

Figure 5. Correlation plots between Chad weather data and NDVI. Note the potential out of phase influence of features manifest in elliptic shapes.

our model can be used for helping those in need, we would discourage use of this model for policy decisions without careful consultation with economists and other professionals.

6. Conclusion

Although understanding the influence of a changing climate on agriculture is an important problem, it remains a challenge. Our work demonstrates the potential for weather forecasting to predict and mitigate the agricultural damages of climate change, especially in developing countries. In the short term, our ARMA models exhibit impressive accuracy in weather prediction. Their quick training time and adaptability make them suitable for model for generalization to various weather patterns in diverse regions. Furthermore, our nonlinear least squares model shows promise as a technique to estimate vegetation growth from seasonal trends. However, our approach also demonstrates the need for a more rigorous treatment of anomalous weather patterns and their effects on local food price data. Better techniques for modeling NDVI as a function of both seasonal trends and extreme weather events could provide useful insight into how plant growth will change in a changing climate.

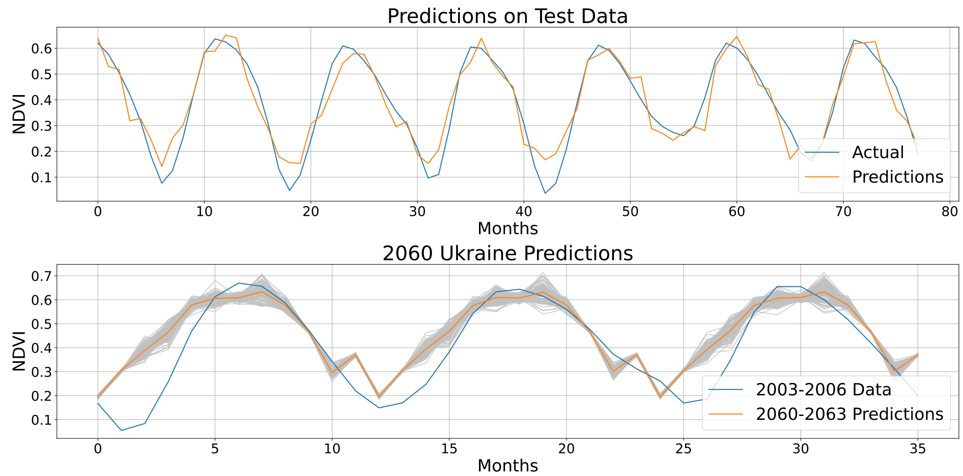

Figure 6. On top we show predictions on the test set for Ukraine compared with actual results suggesting our model is able to generally predict NDVI from weather data. In particular, the second to last full trough in the prediction follows the elevated trough in the actual data but still drops down for the last full trough, showing that it is able to capture variations in the trend. On the bottom we plot NDVI predictions for generated weather data for the year 2060. Sampled NDVI trends are shown in grey

References

[Ado24] Philip Kofi Adom. The socioeconomic impact of climate change in developing countries over the next decades: A literature survey. *Heliyon*, 10(15):e35134, July 2024. Accessed: 2025-04-12.

[AKA+24] Muhammad Ashfaq, Imran Khan, Abdulrahman Alzahrani, Muhammad Usman Tariq, Humera Khan, and Anwar Ghani. Accurate wheat yield prediction using machine learning and climate-ndvi data fusion. *IEEE Access*, 12:40947–40961, 2024.

[BK15] Molly E. Brown and Varun Kshirsagar. Weather and international price shocks on food prices in the developing world. *Global Environmental Change*, 35:31–40, 2015.

[Geo] GIS Geography. What is ndvi (normalized difference vegetation index)?

[GIS25] GISGeography. Ndvi (normalized difference vegetation index), 2025. Retrieved March 31, 2025.

[NOA25] NOAA National Centers for Environmental Information. 2024 was warmest year in the modern record for the globe, January 2025. Accessed: 2025-04-12.

[oA25] U.S. Department of Agriculture. Crop explorer: Southeast asia legend description. U.S. Department of Agriculture, Foreign Agricultural Service, 2025. Retrieved March 31, 2025.

[Unk17] Unknown. Forecasting wheat yield from meteorological data and MODIS NDVI data across the punjab province, Pakistan. *International Journal of Remote Sensing*, Unknown(Unknown):Unknown, 2017.

[Wor25] World Bank. Ukraine – trends variability – projections, 2025.