

Ronan Lucey (ronan.m.lucey@nasa.gov)¹, W. Lee Ellenburg^{1,2}, Robert Griffin^{1,2}, Udaysankar Nair¹

(1)University of Alabama in Huntsville Earth System Science Center, 320 Sparkman Drive NW, Huntsville, AL 35805 (2) NASA SERVIR Science Coordination Office, 320 Sparkman Drive NW, Huntsville, AL 35805

Overview

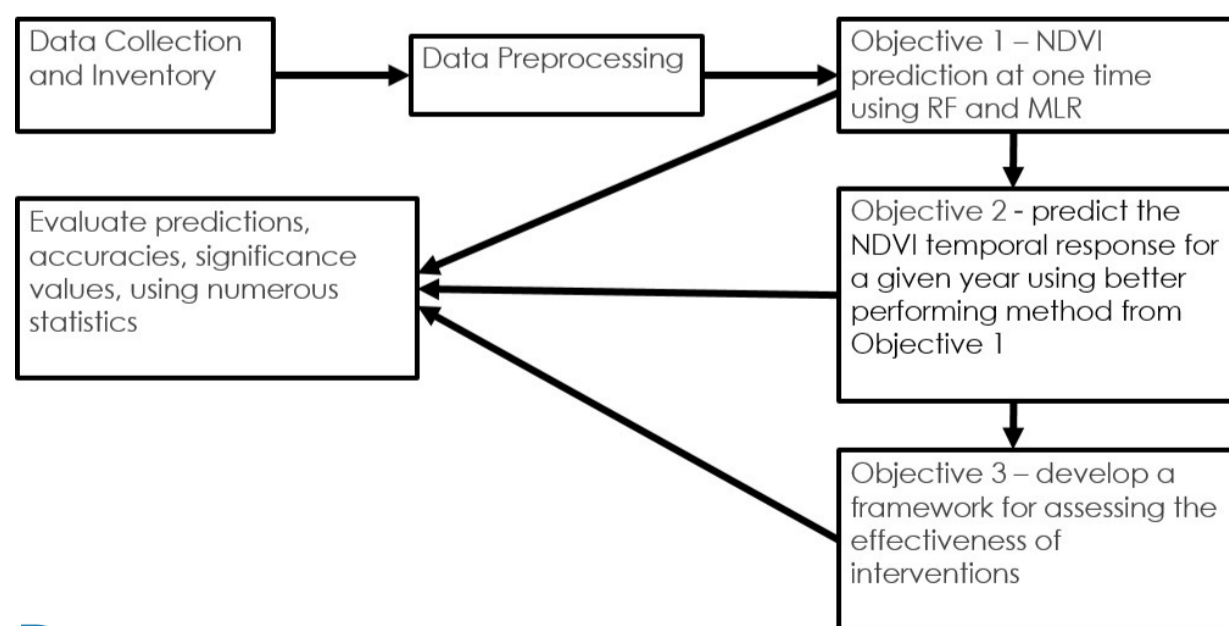
Food security in Nepal is a major issue in the face of a changing climate. Agricultural interventions, which foster the modernization of agricultural systems by deliberate and strategic actions on the part of governments or organizations, are being performed with the main goals of increasing production, profit, and nutrition for the poor. Examples of agricultural interventions include irrigation, machinery, genetically modified seeds, advanced weather reports, and access to outside resources such as resource workers. However, assessment of the effectiveness of agricultural interventions can be difficult for many reasons, including cost, time needed, and the remoteness of the Nepalese countryside.

This research is an effort to assess a new impact evaluation approach that utilizes remote sensing data, addresses the challenges of traditional assessment methods, and builds off other ongoing initiatives in the region. We use the Normalized Difference Vegetation Index (NDVI) as a proxy for production. Production is a signal from agriculture that can be connected to satellites through indices such as NDVI. The overarching purpose of the research is to compare a counterfactual predicted production to the observed actual production from satellites, and whether there is a statistically significant difference that can be observed between them.

Research Questions

- U** Can we accurately predict NDVI at a single time using remotely sensed data and multi-linear regression (MLR) or random forest (RF) techniques?
- V** Can we accurately predict the temporal response of a NDVI time series for a given year?
- W** Can we develop a framework for assessing the impact of agricultural interventions using remotely sensed data and the selected statistical technique?

Methodology

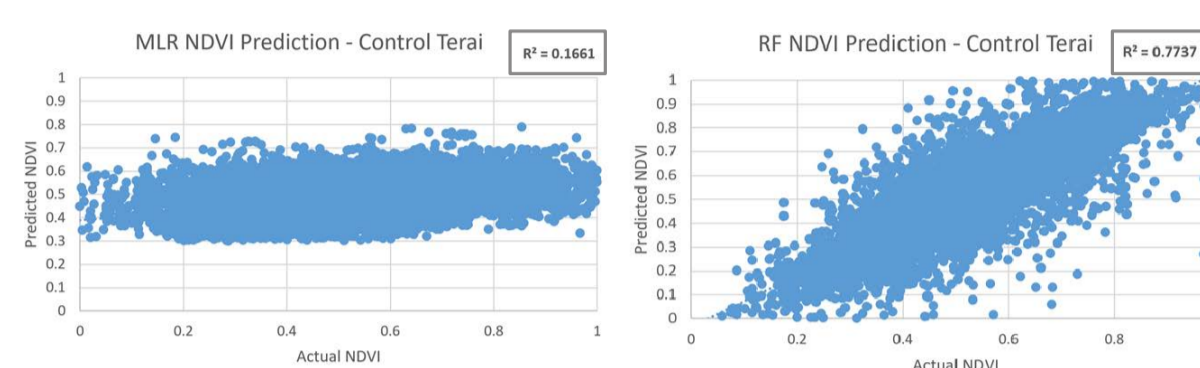


Datasets



* South Asia Land Data Assimilation System (SALDAS) - Ben Zaitchik, Johns Hopkins University

Results

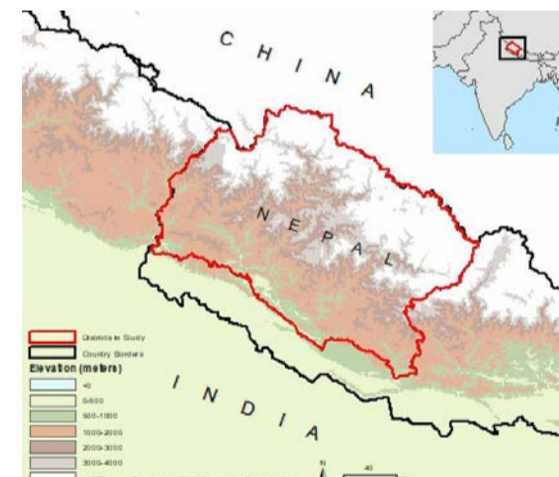


Model	Mean Adjusted R^2	Mean RMSE	Mean MAE
RF Models	0.706	0.101	0.068
MLR Models	0.206	0.163	0.124

- Two approaches tested for their ability to predict NDVI at a single time using remotely sensed data - **multi-linear regression (MLR) approach** (left figure) and **random forest (RF) approach** (right figure)
- Tests done over three different elevation zones in Nepal - the low elevation Terai region (<1200 m), the hilly Siwalik region (1200-2400 m), and the High Mountain region (>2400 m)
- **RF models outperformed the MLR models at all elevation zones, with R^2 values averaging 0.706 across all zones** (above table)
- As a result, only the RF models were used in addressing research questions 2 and 3, and MLR was not used further

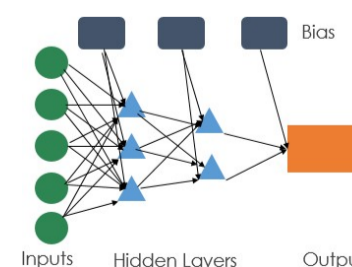
Study Area

- The study area for this research included **22 districts in the West and Far-West regions of Nepal** (right figure). ~1900 fields from area were used
- Nepal is a landlocked country in the Hindu-Kush Himalaya region with a steep elevation gradient and many climate zones

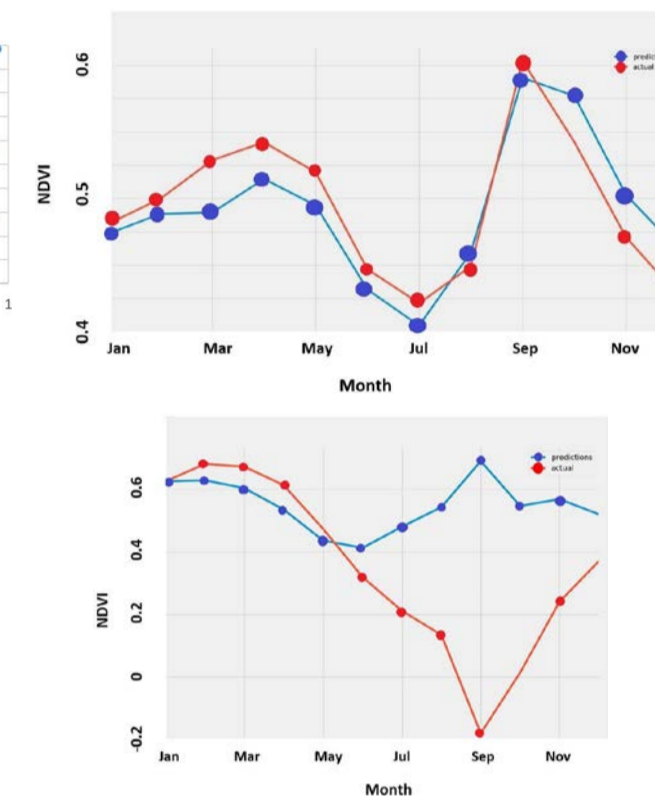


Future Directions

- Test using **in-situ data as the dependent variable for production**, instead of using NDVI as a proxy
- Include **additional independent variables** (soil composition, soil moisture, known crop type/growing patterns)
- Testing this methodology over **different areas of the world**, Testing this methodology on **different years**, both amount of input data and amount of time to predict on
- A **machine learning approach using an Artificial Neural Network (ANN)** could be optimized and potentially improve on RF/MLR approaches
- An ANN (right figure) finds the weights that maximize accuracy to be given to each input by **back-propagating error**



V



- Validation using RF technique of NDVI time series prediction was conducted
- **Model was trained on fields that were not intervened on and tested on fields that were not intervened on by back-predicting one year of NDVI**
- The accuracy for the resulting predicted NDVI time series was **high for some fields (top figure) but low for others (bottom figure)**
- The overall accuracy for time series NDVI prediction across all fields was lower than that found predicting a single time, regardless of stratification (table below)

Random Forest Model(s)	Mean Adjusted R^2
Control	0.312
Delayed	0.348
Early	0.465
Terai	0.408
Siwalik	0.474
Mountain	0.244

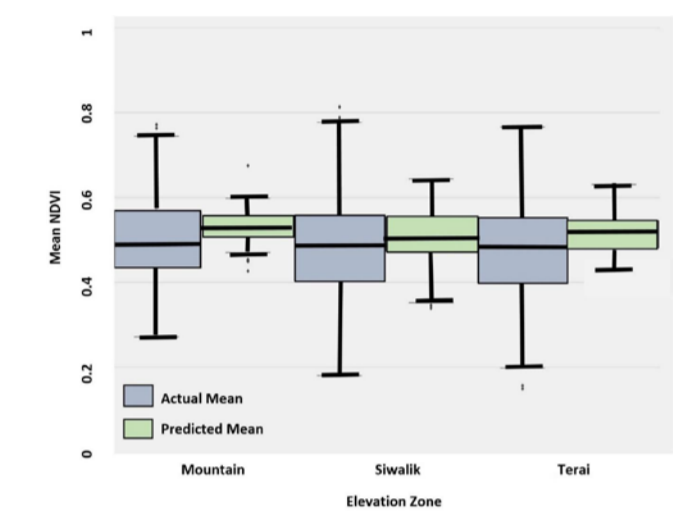
Impact / Conclusions

- This research **responds to a need** identified by SERVIR-HKH for effective ways to assess agricultural productivity, and supports the agricultural efforts to combat food insecurity ongoing in the region
- The RF approach was **able to predict single-time NDVI values** with the desired level of accuracy, unlike the MLR approach, but was **unable to accurately predict a time series of NDVI values** that reflected the observed values via satellite
- Despite the lack of confidence in our results for the third research question, **a framework has been developed that can be utilized for agricultural intervention assessment in the future**

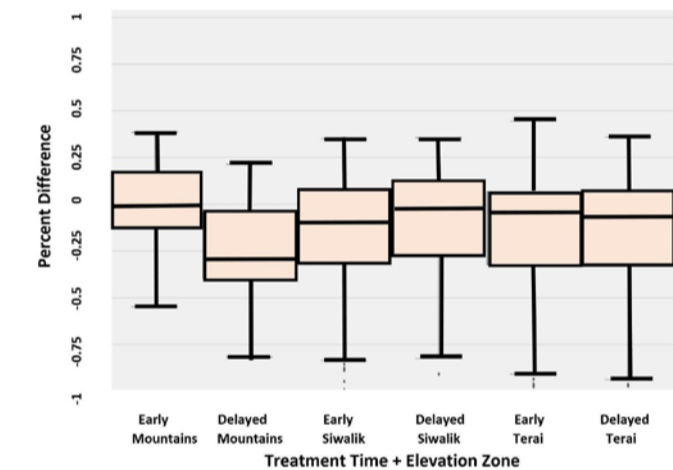
Acknowledgement

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration (NASA). Any mention of commercial product, service, or activity in this material does not constitute NASA endorsement. This material is based upon work supported by NASA through cooperative agreement NNM11AA01A.

W



- The model trained on control fields was utilized for NDVI prediction over fields that had been treated
- **The mean annual NDVI of tested fields, stratified by elevation, was higher for the counterfactual NDVI than for the observed NDVI** (above figure)



- **Percent difference between predicted control and observed treated NDVI values for each field was calculated for each elevation zone and treatment time** (above figure)
- **Low level of confidence with the results (P Value = 0.86)**, but framework for intervention assessment using remote sensing and statistical techniques was **successfully developed and tested**.