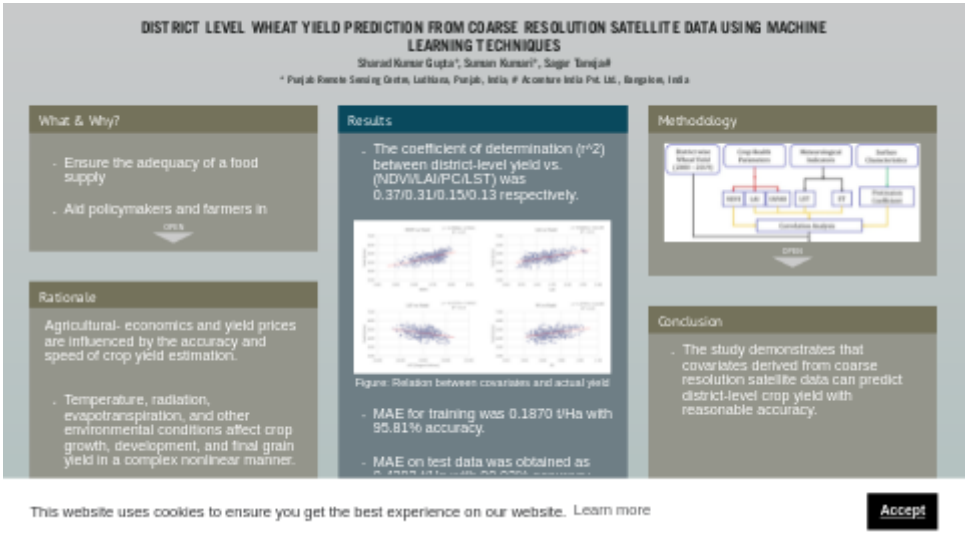
DISTRICT LEVEL WHEAT YIELD PREDICTION FROM COARSE RESOLUTION SATELLITE DATA USING MACHINE LEARNING TECHNIQUES



Sharad Kumar Gupta, Suman Kumari, Sagar Taneja

Punjab Remote Sensing Centre, Ludhiana, Punjab, India

PRESENTED AT:

WHAT & WHY?

- Ensure the adequacy of a food supply
- Aid policymakers and farmers in managing harvest, import/ export, & transportation
- Anticipate market fluctuations
- Yield forecast plays a critical role in decision-making for reducing the yield gap.

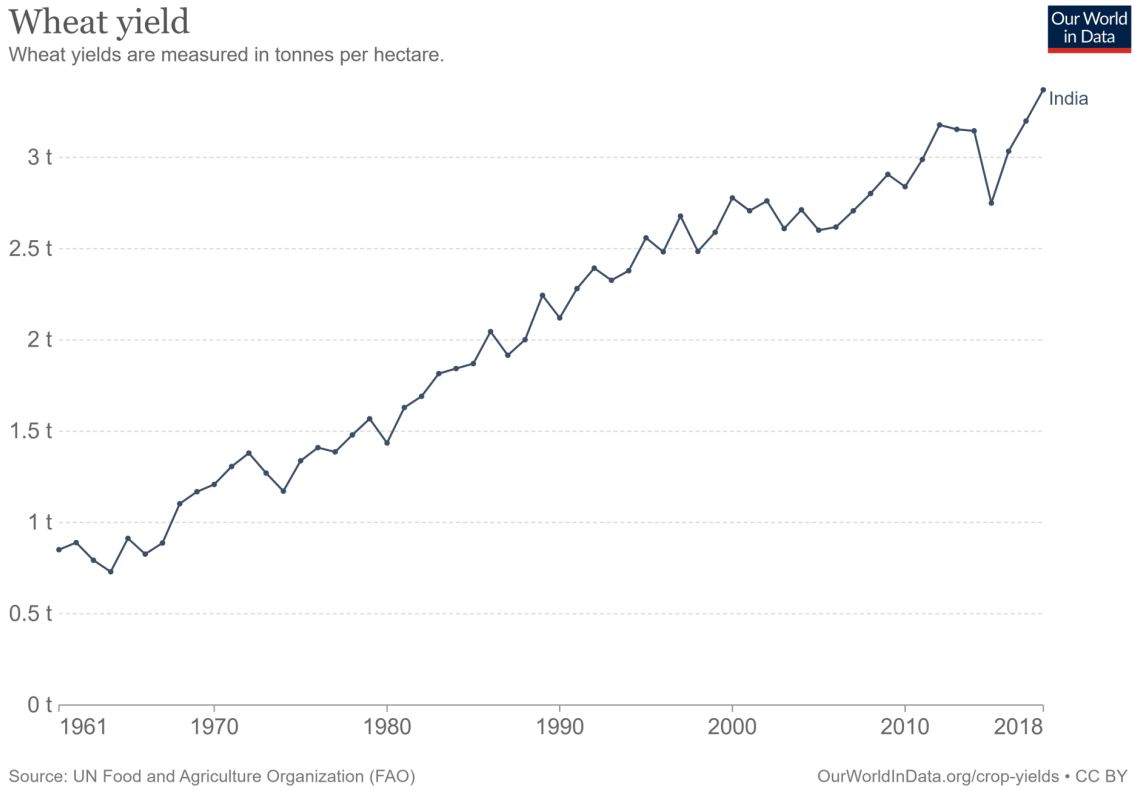


Figure: Wheat yields measured in tonnes per hectare (Source: Our World in Data, FAO)

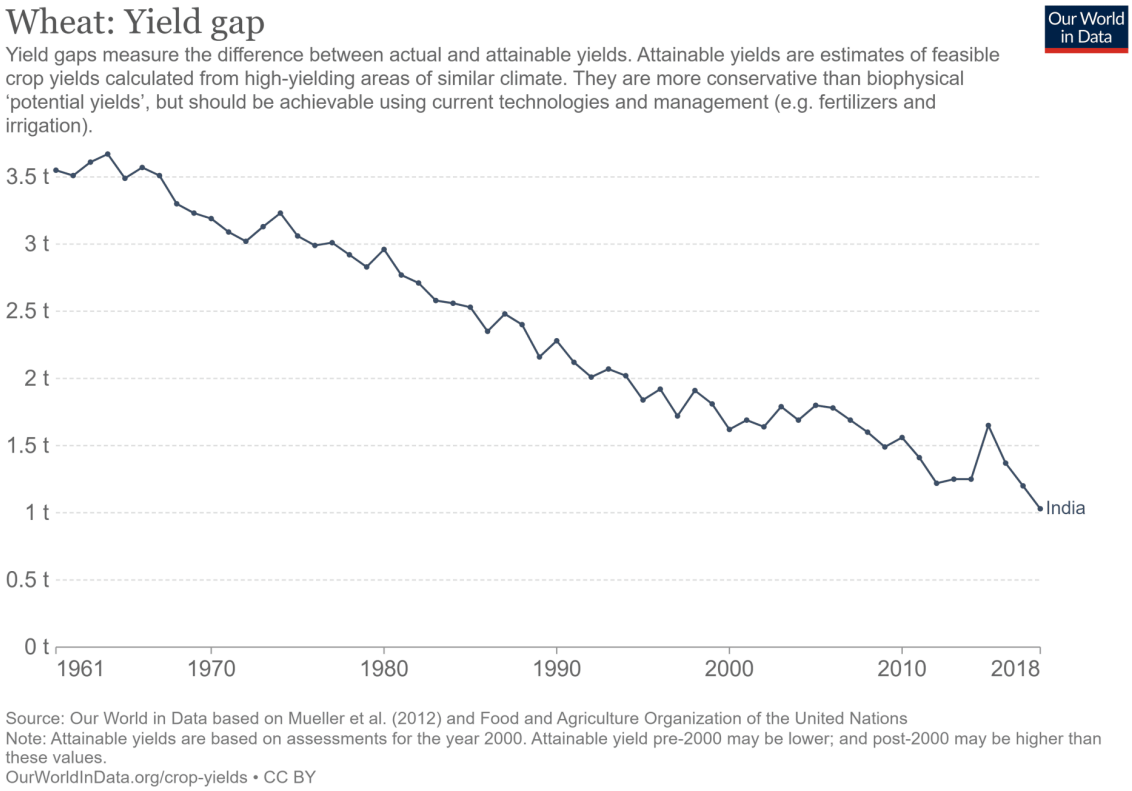


Figure: Yield gaps measure the difference between actual and attainable yields. (Source: Our World in Data, FAO)

Note: Attainable yields are estimates of feasible crop yields calculated from high-yielding areas of similar climate. They are more conservative than biophysical 'potential yields', but should be achievable using current technologies and management (e.g. fertilizers and irrigation).

RATIONALE

Agricultural- economics and yield prices are influenced by the accuracy and speed of crop yield estimation.

- Temperature, radiation, evapotranspiration, and other environmental conditions affect crop growth, development, and final grain yield in a complex nonlinear manner.

- Machine learning (ML) techniques can account for such nonlinear relations between yield and its covariates.

RESULTS

- The coefficient of determination (r^2) between district-level yield vs. (NDVI/LAI/PC/LST) was 0.37/0.31/0.15/0.13 respectively.

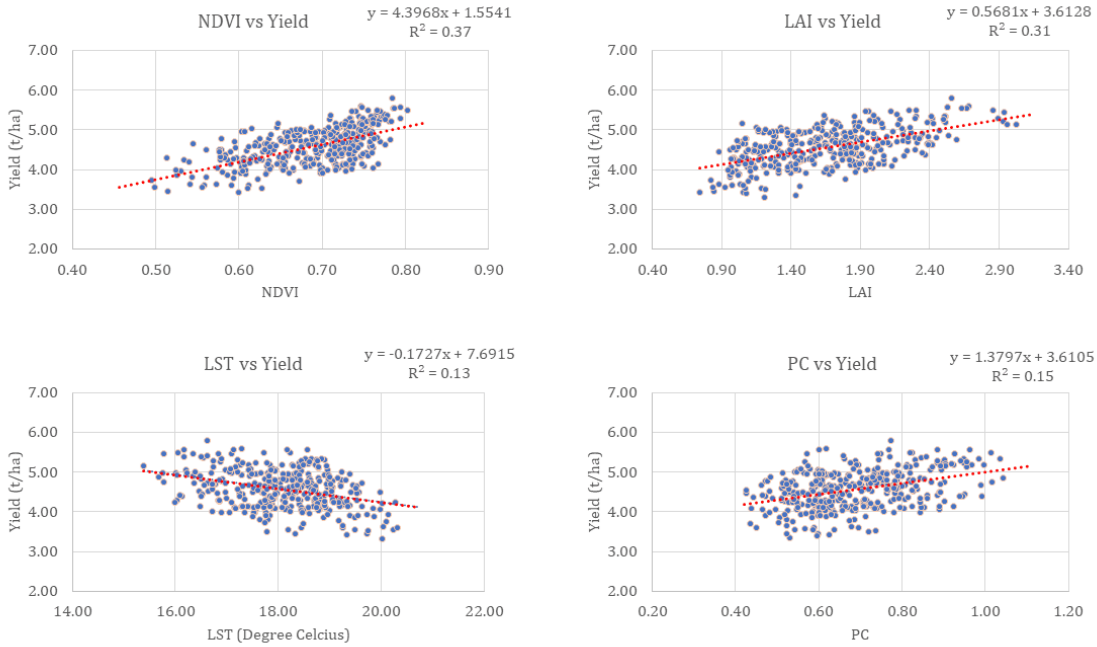


Figure: Relation between covariates and district level yield

- MAE for training was 0.1870 t/Ha with 95.81% accuracy.

- MAE on test data was obtained as 0.4293 t/Ha with 90.02% accuracy.

- The results are within acceptable error limits of the published research articles.

METHODOLOGY

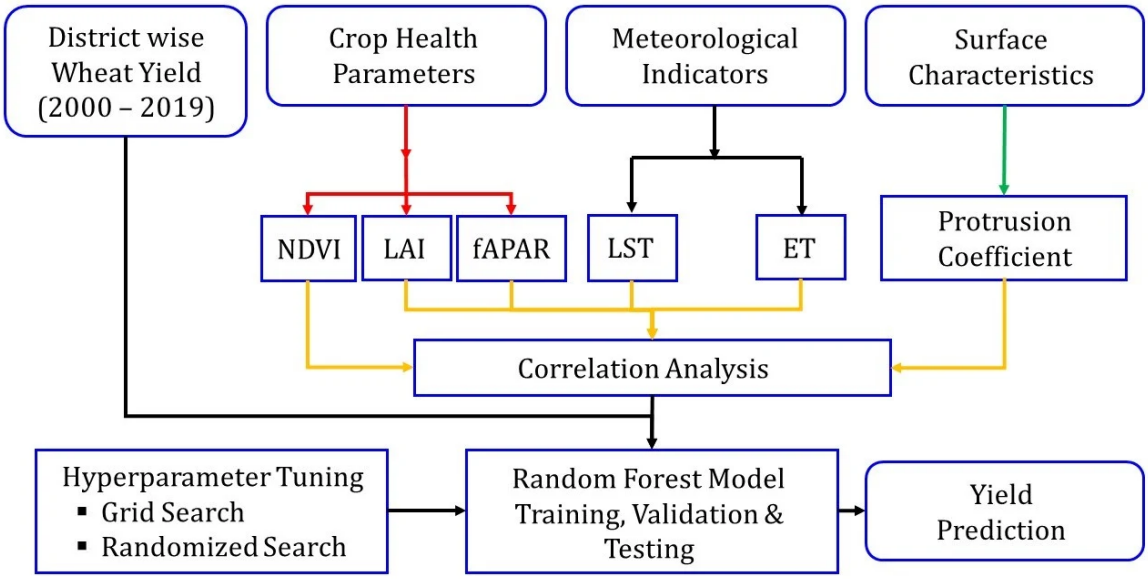


Figure: Flowchart of the methodology adopted in this study

- Wheat yield estimates were available for the year 2000 – 2019.
- Covariates for crop health such as normalized difference vegetation index (NDVI), leaf area index (LAI), the fraction of absorbed photosynthetically active radiation (fAPAR).
- Meteorological indicators such as land surface temperature (LST), and evapotranspiration (ET).
- Surface characteristics such as protrusion coefficient (PC).
- Products generated at 250 m spatial resolution from the MODIS data using Google Earth Engine.
- Training (2000 – 2009, 2011, 2013, 2014, 2016 - 2019)

- Testing (2010, 2012, 2015)

- Used the random forest (RF) regression method to create a wheat yield prediction model

CONCLUSION

- The study demonstrates that covariates derived from coarse resolution satellite data can predict district-level crop yield with reasonable accuracy.

ABSTRACT

Regional crop production estimates are important in both public and private sectors to ensure the adequacy of a food supply and aid policymakers and farmers in managing harvest, storage, import/export, transportation, and anticipate market fluctuations. Food security will be progressively challenged by population growth and climate change. Thus, the prediction of accurate regional crop yield is essential for national food security and the sustainable development of the Indian agriculture sector.

In this study, we have selected Punjab, the highest wheat yielding state in India. The district-wise wheat yield data were available for the year 2000 – 2019. We have used several covariates for crop health viz. normalized difference vegetation index (NDVI), leaf area index (LAI), fraction of absorbed photosynthetically active radiation (fAPAR); meteorological indicators viz. land surface temperature (LST), and evapotranspiration (ET); and surface characteristics viz. protrusion coefficient (PC). These indicators were generated at 250 m spatial resolution from the MODIS data using Google Earth Engine. The whole data was divided into two groups for training (2000 – 2009, 2011, 2013, 2014, 2016 - 2019) and testing (2010, 2012, 2015), which were randomly selected. This study uses the random forest (RF) regression method to create a wheat yield prediction model. We created several combinations of covariates and found that fAPAR and ET are highly correlated with NDVI and do not have much influence on the model's prediction accuracy. Hence, only four out of six covariates were selected for final training. The coefficient of determination between district-level yield vs. (NDVI/LAI/PC/LST) was 0.37/0.31/0.15/0.13 respectively. We used randomized search cross-validation as well as grid search cross-validation for hyper-parameter tuning. Furthermore, we used mean absolute error (MAE) and accuracy as quality metrics. The MAE for training was 0.1870 t/Ha with 95.81% accuracy, whereas the MAE on test data was obtained as 0.4293 t/Ha with 90.02% accuracy. The results of this study are within acceptable error limits of the published research articles. Overall, this study demonstrates that covariates derived from coarse resolution satellite data can predict district-level crop yield with reasonable accuracy.