# MODELLING BIOPHYSICAL PARAMETERS OF MAIZE USING LANDSAT 8 TIME SERIES


**T. Dahms[1], S. Seissiger[1], E. Borg[2], C. Conrad[1]**

[1] University of Würzburg, Department of Geography and Geology, Department of Remote Sensing
Oswald-Külpe-Weg 86, 97074 Würzburg, Germany

[2] German Aerospace Center (DLR), National Ground Segment,
Kalkhorstweg 53, 17235 Neustrelitz, Germany

Contact: thorsten.dahms@uni-wuerzburg.de


## Aim

The study aims at an optimized prediction of biophysical parameters of maize and the identification of the best explaining spectral bands and vegetation indices from Landsat 8 OLI sensor. For this purpose, we used an in-situ dataset from 06.05.2015 to 15.10.2015. Random forest and Conditional Inference Forests were used because of their explicit strong exploratory and predictive character. Variable importance measures allowed for analysing the relation between the biophysical and the spectral response, and the performance of the two approaches over the plant stock evolvement of maize.

## Study Site

This study was implemented on the TERENO test site DEMMIN in Mecklenburg-Western Pomerania (Figure 1). A field campaign was conducted in collaboration with the calibration and validation facility DEMMIN (DLR) during the vegetation period 2015. The field observations were carried out on 18 Environmental Sampling Units (ESUs) between May and October 2015 in a weekly to bi-weekly tonus (Figure 3).

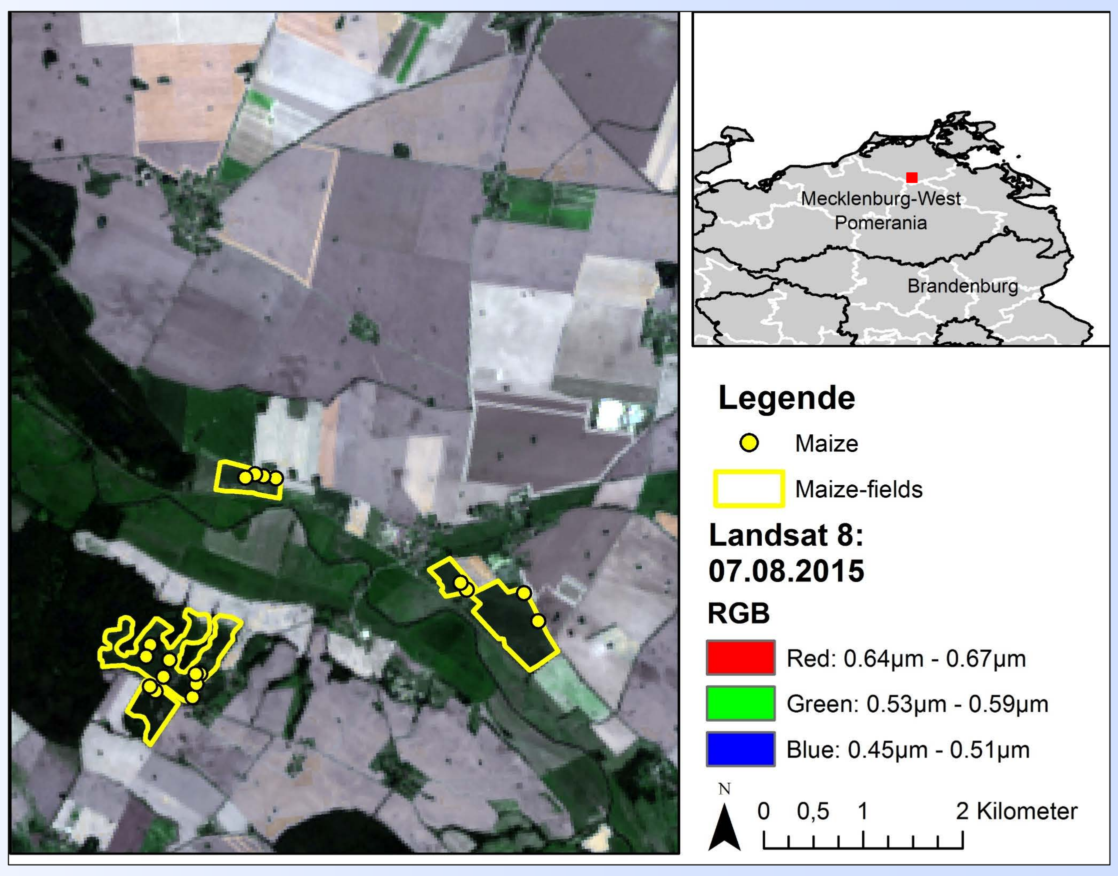


Figure 1: Distribution of the ESUs on five maize fields in the study area

## Data and Methods

### Ground Observation

The Fraction of Photosynthetically Active Radiation (FPAR), the Leaf Area Index (LAI) and the Chlorophyll Content (SPAD) were repeatedly measured on twelve points of one ESU (e.g. Figure: 2). These twelve measurements of the respective biophysical parameter were averaged. The averages were later on used in the machine learning methods as response variable.

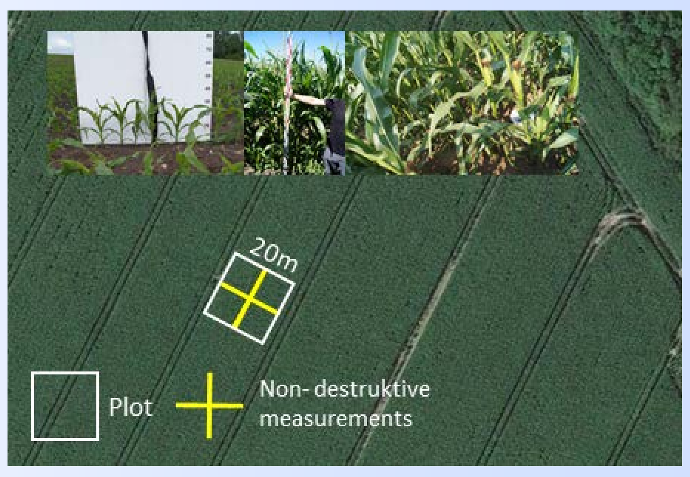


Figure 2: Sampling of FPAR, LAI and SPAD in maize field

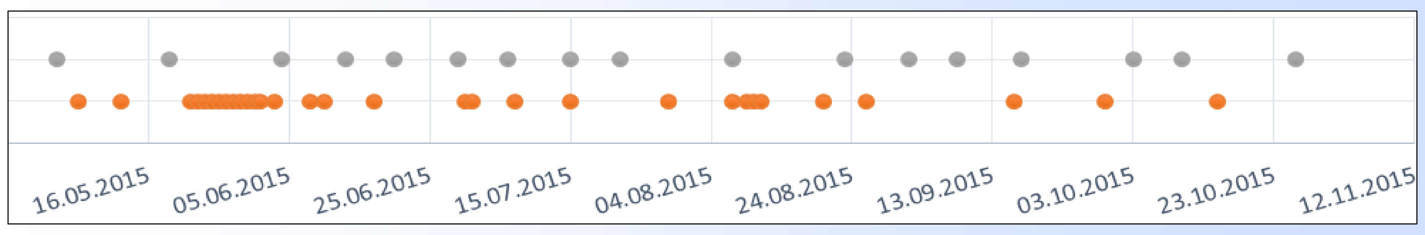


Figure 3: Landsat 8 OLI and ground observations over the vegetation period of 2015 for maize

- 17 Landsat 8 OLI Scenes
- Weekly to bi-weekly ground observation

### Remote Sensing Dataset

17 Landsat 8 OLI scenes were atmospherically corrected and cloud masked. In addition to the seven OLI bands (Costal, Blue, Green, Red, NIR, SWIR_1, SWIR_2), five Vegetation Indices (NDVI, SAVI, RDVI, SR, EVI) as well as six Tasseled Cap Indices (Brightness, Greenness Wetness, TCT4, TCT5, TCT6) were calculated. This Index-Band ensemble was used as predictor dataset in the machine learning models.

### Modelling Biophysical Parameters

We used Random Forest (rforest) [1] and Conditional Inference Forest (cforest) [3] to model biophysical parameters, namely FPAR, LAI and SPAD on maize for the entire vegetation period. In contrast to the rforest, cforest is built from conditional inference trees which are able to consider cause-effect relations during variable selection and to reduce bias in case of highly correlated variables. A similar procedure was developed for an unbiased extraction of variable importance. It identifies those variables which mostly influence the accuracy in the regression tree ensemble. Both methods were executed 100 times. Every rforest or cforest was tuned using ten different mtry values (number of variables considered for each split) (2,3,5,7,9,10,12,14,16,18 with p=18), while the number of trees was held fix at 500. The so called cforest routine and the conditional variable importance algorithm are implemented in the party package [2] in the statistic software R. Each run was validated using a five fold cross validation. The accuracy and the variable importance were extracted after each run. The distribution of the variable importance of rforest and cforest can be seen in Figure 4, while the accuracies are displayed in Table 1.

## Results

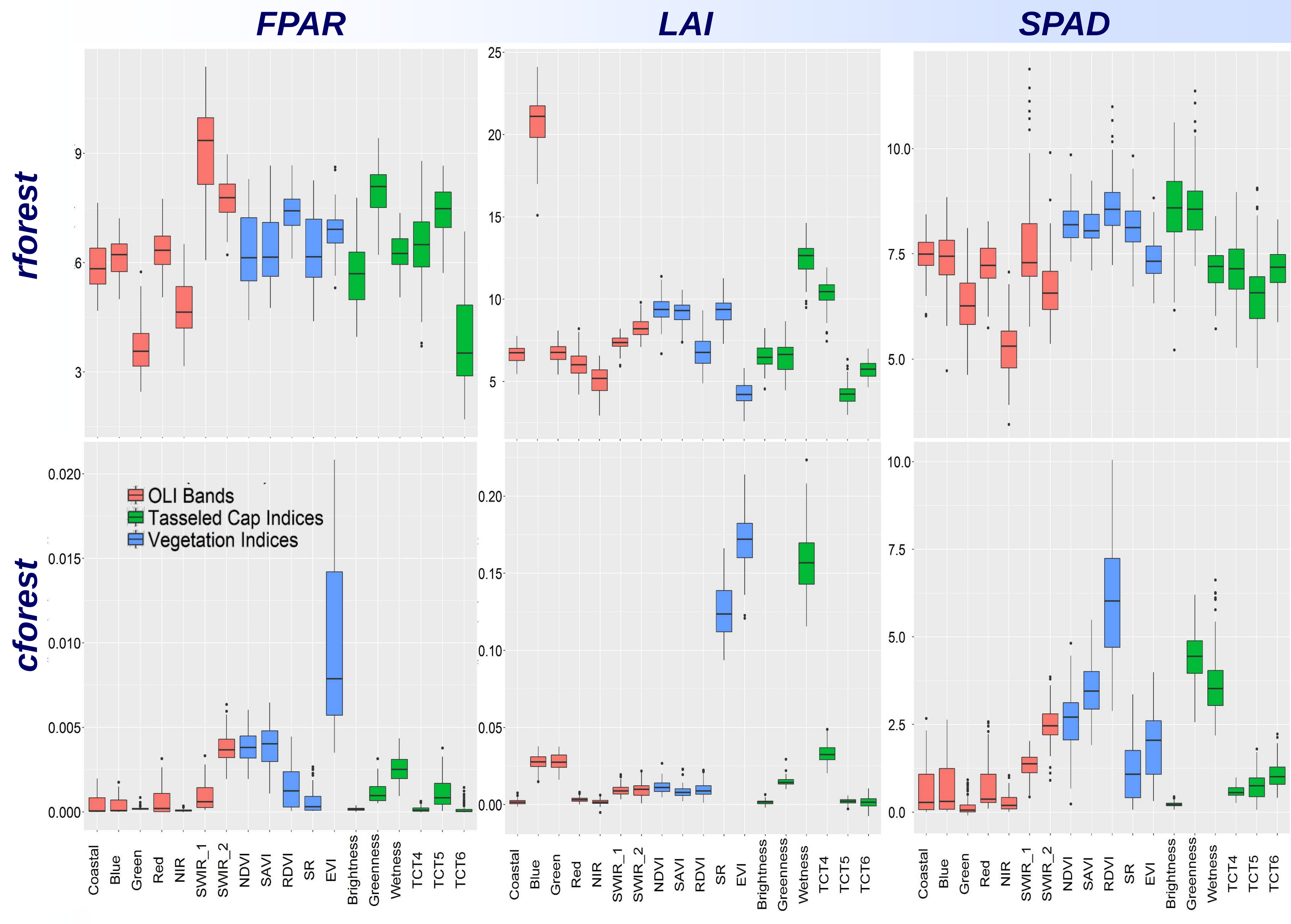


Figure 4: Variable importance distribution of the predictor variables (Landsat OLI Band + Vegetation Indices + Tasseled Cap Indices) of Random Forest (rforest) and Conditional Inference Forests (cforest) modelling FPAR, LAI and SPAD for the entire vegetation period of maize 2015.

| | | FPAR | LAI | SPAD |
|---|---|---|---|---|
| rforest | $R^2$ | 0.85 | 0.70 | 0.83 |
| | RMSE | 0.11 | 0.8 | 4.5 |
| | mtry | 15 | 10 | 2 |
| cforest | $R^2$ | 0.85 | 0.64 | 0.80 |
| | RMSE | 0.11 | 0.9 | 4.9 |
| | mtry | 3 | 13 | 10 |
| | Samples | 93 | 92 | 94 |

Table 1: Comparison between the mean model performance of Random Forest (rforest) and Conditional Inference Forest (cforest) over 100 runs. Additionally showing the most often chosen mtry value and the sample size are shown.

## Conclusion and Remarks

The variable importance boxplots of the Conditional Inference Forest models show a clearer distribution than the boxplots of the Random Forest models. Vegetation Indices (esp. EVI) seem to be the most suited predictors modelling biophysical parameters using Landsat OLI data. The Random Forest trends to rank the single OLI Bands higher than the Vegetation Indices or the Tasseled Cap Indices. For FPAR and SPAD, Random Forest showed a very vague ranking in the importance of the predictor variables. For modelling the LAI the Random Forest shows an outstanding importance of the Blue band. In terms of model performance, Random Forest models equal or outrun the Conditional Inference Forest for every biophysical parameter.

Altogether, the study showed that the Random Forest has higher predictive power but smaller explorative character than the Conditional Inference Forest.

## References


[1] Breiman, L. (2001). Random forests. Machine learning45(1), 5-32.

[2] Horthorn, T., Hornik, K., Strobl, C., Zeileis, A., 2015: A Laboratory for Recursive Partytioning

[3] Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. BMC bioinformatics, 9(1), 1.



Techs4TimeS is funded by:
Bundesministerium für Wirtschaft und Energie
FKZ No. 50 EE1353