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 evolution of maize.

**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.

**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 forest 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.

**Conclusion and Remarks**

The variable importance boxplots of the Conditional Inference Forest models show a clearer distribution than the Random Forest models. Vegetation Indices (esp. EVI) seem to be the most suited predictors for 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 models equal or outrun the Conditional Inference Forest.

**Results**

The 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).

**References**

