# IMPORTANT VARIABLES MODELLING BIOPHYSICAL PARAMETERS ON WINTER WHEAT

**T. Dahms[1], S. Seissiger[1], E. Borg[2], H.Vajen[2], B. Fichtelmann[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 winter wheat and the identification of the best explaining spectral bands and vegetation indices from the RapidEye sensor system. For this purpose, we used an in-situ dataset of biophysical parameters from 24.03.2015 to 07.08.2015. Conditional Inference Random Forests [2] (Cforest) were used because of their explicit strong exploratory character. Variable importance measures allowed for analysing the relation between the biophysical and the spectral response. The performance of the Cforest was analysed using the $R^2$ and the RMSE values.

## 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 March and August 2015 in a weekly to bi-weekly tonus (Figure 3).

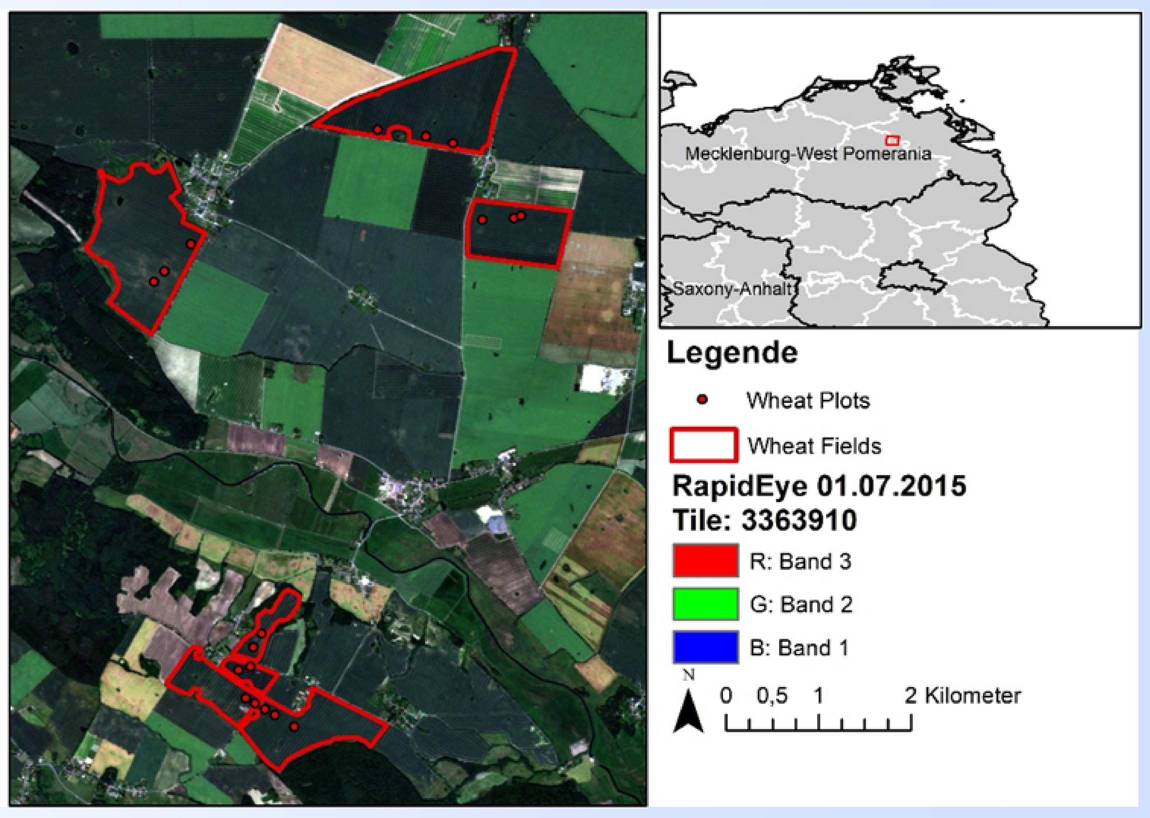



Figure 1: Distribution of the ESUs on seven winter wheat fields in the study area

## Data and Methods

### Ground Observation

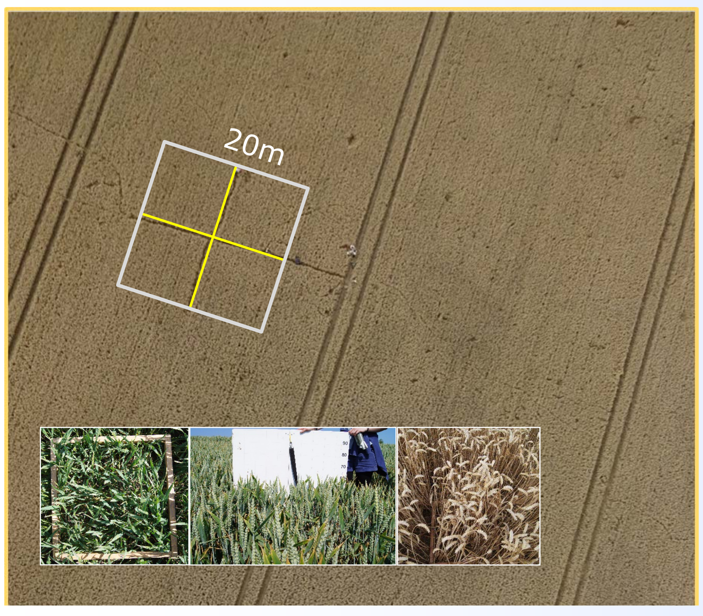



Figure 2: Sampling of FPAR and LAI in a winter wheat field

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 Cforest as response variable.

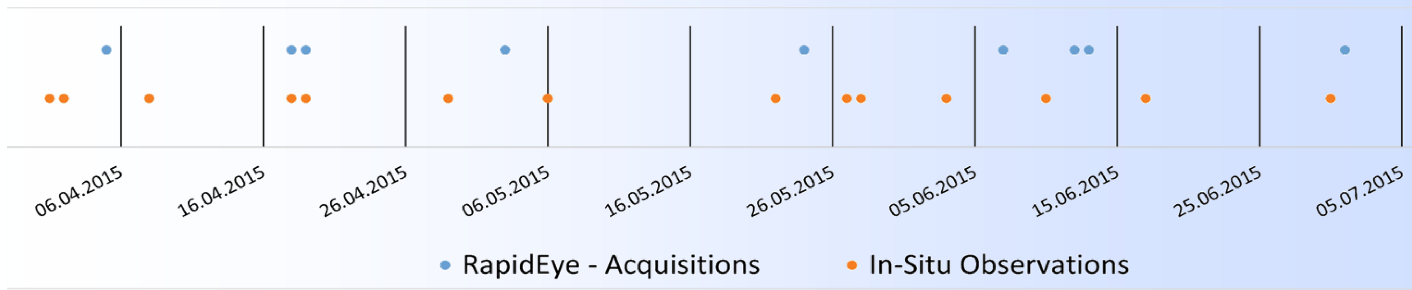


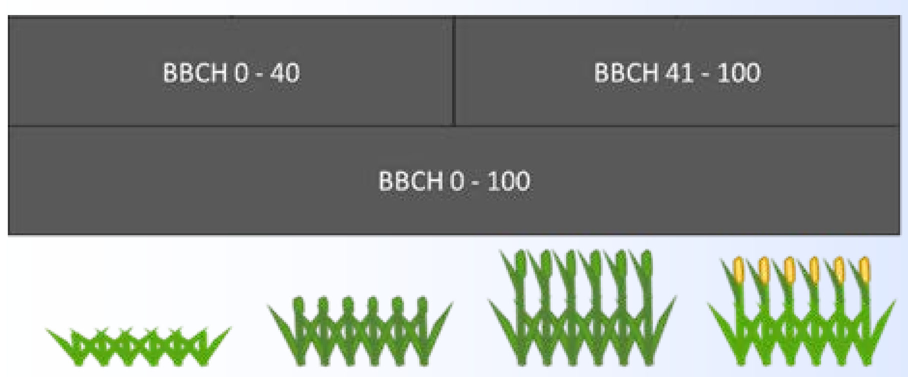

Figure 3: RapidEye and in-situ observation over the vegetation period of 2015 for winter wheat

### Remote Sensing Dataset

Nine RapidEye scenes were atmospherically corrected and cloud masked. In addition to the five RapidEye bands (Blue, Green, Red, RedEdge) nine Vegetation Indices (SR, NDVI, SAVI, RE_NDVI, RDVI, EVI, Curv, Length and relLength) as well as three Tasseled Cap Indices (Brightness, Greenness, Wetness) were calculated. This Index-Band ensemble was used as predictor dataset in the Cforest.

### Modelling Biophysical Parameters

We used the Conditional Inference Random Forest (Cforest) to model biophysical parameters on winter wheat, namely FPAR, LAI and SPAD. Doing so, we investigated the model performance as well as the variable importance. This investigation was carried out for different stages of the plant evolvement (e.g. Figure 4).




| BBCH Code | Phenological Group |
|---|---|
| 0-40 BBCH | Germination to Booting |
| 41-100 BBCH | Booting to Senescence |
| 0-100 BBCH | All stages |

Figure 4: Phenological groups of the ground observations

Cforest is built from conditional inference trees [1] 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. We run the model 100 times. Every Cforest was tuned using ten different mtry values (number of variables considered for each split) (2,3,4,6,8,9,11,13,15,17 with p=17), while the number of trees was held fix at 500.

## Results

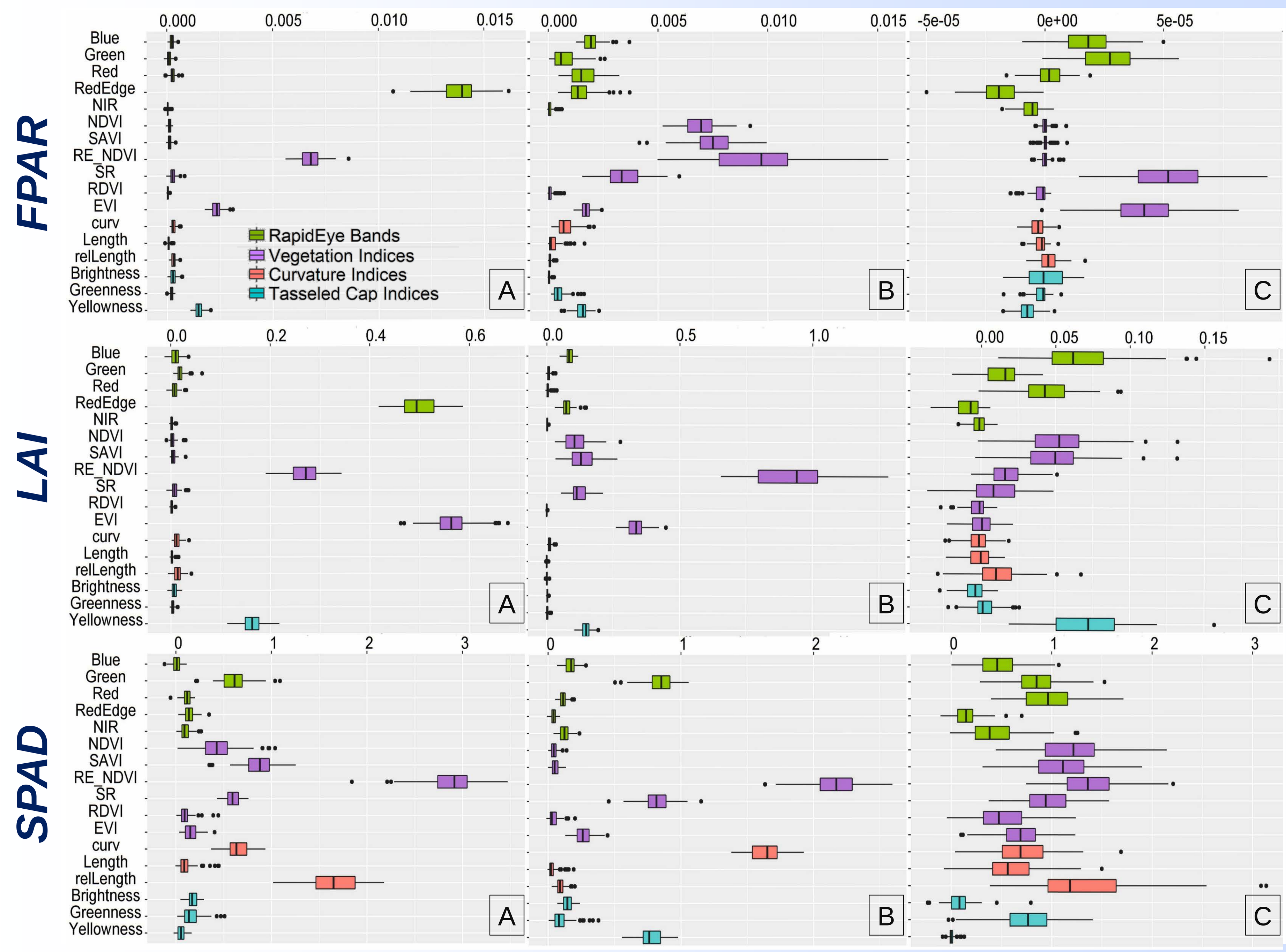



Figure 5: Variable importance distribution boxplots over 100 Cforest runs for the phenological groups A: 0-100 BBCH; B: 0-40 BBCH; C: 41-100 BBCH

| | BBCH | 0-100 | 0-40 | 41-100 |
|---|---|---|---|---|
| FPAR | RMSE | 0.16 | 0.12 | 0.04 |
| | $R^2$ | 0.59 | 0.83 | 0.21 |
| | mtry | 12 | 5 | 17 |
| | samples | 124 | 68 | 56 |
| LAI | RMSE | 1.56 | 1.23 | 1.87 |
| | $R^2$ | 0.41 | 0.66 | 0.33 |
| | mtry | 12 | 10 | 2 |
| | samples | 111 | 62 | 49 |
| SPAD | RMSE | 4.98 | 3.17 | 6.94 |
| | $R^2$ | 0.29 | 0.42 | 0.28 |
| | mtry | 12 | 17 | 2 |
| | samples | 161 | 103 | 58 |

Table 1: Comparison between the mean model performance of Conditional Inference Forest (Cforest) over 100 runs for different phenological groups. Additionally showing the most often chosen mtry value and the sample size.

## Conclusion and Remarks

The variable importance boxplots of the Cforest models show a clear distribution for the entire vegetation period (A) and for the growing period (B), while the distribution seems to be more vague for the senescence (C). Vegetation Indices (esp. RE_NDVI) appear to be the most suited predictors modelling biophysical parameters using RapidEye. The RedEdge band seems to be very important modelling FPAR and LAI for the entire vegetation period. Table 1 illustrates, that the model performance varies between the different phenological groups. The $R^2$ values for the growing stages are always higher than for the senescence.

Altogether, the study showed that the RedEdge and the RedEdge based vegetation index RE_NDVI are very important for modelling biophysical parameter in the growing stages. Moreover it also demonstrates, that the model performance decreases and the variable importance gets more vague in the time of senescence.

## References

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

[2] 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