Sentinel-2 based tree type classification in Bavaria using imbalanced datasets

S. Förtsch*, M. Thiel*, I. Otte*, T. Ullmann*
* Department of Remote Sensing, University of Wuerzburg

1. Introduction

The launch of ESAs (European Space Agency) Multispectral Instrument (MSI) aboard Sentinel-2A and Sentinel-2B increases the availability of free satellite data in the optical domain significantly. This is also affecting the quality of tree type classifications.
However, the lack of appropriate ground truth is still a bottleneck to unlock the full potential of satellite data. Therefore we examine the methods of building models on regional scale and imbalanced ground truth for classifying larger areas.

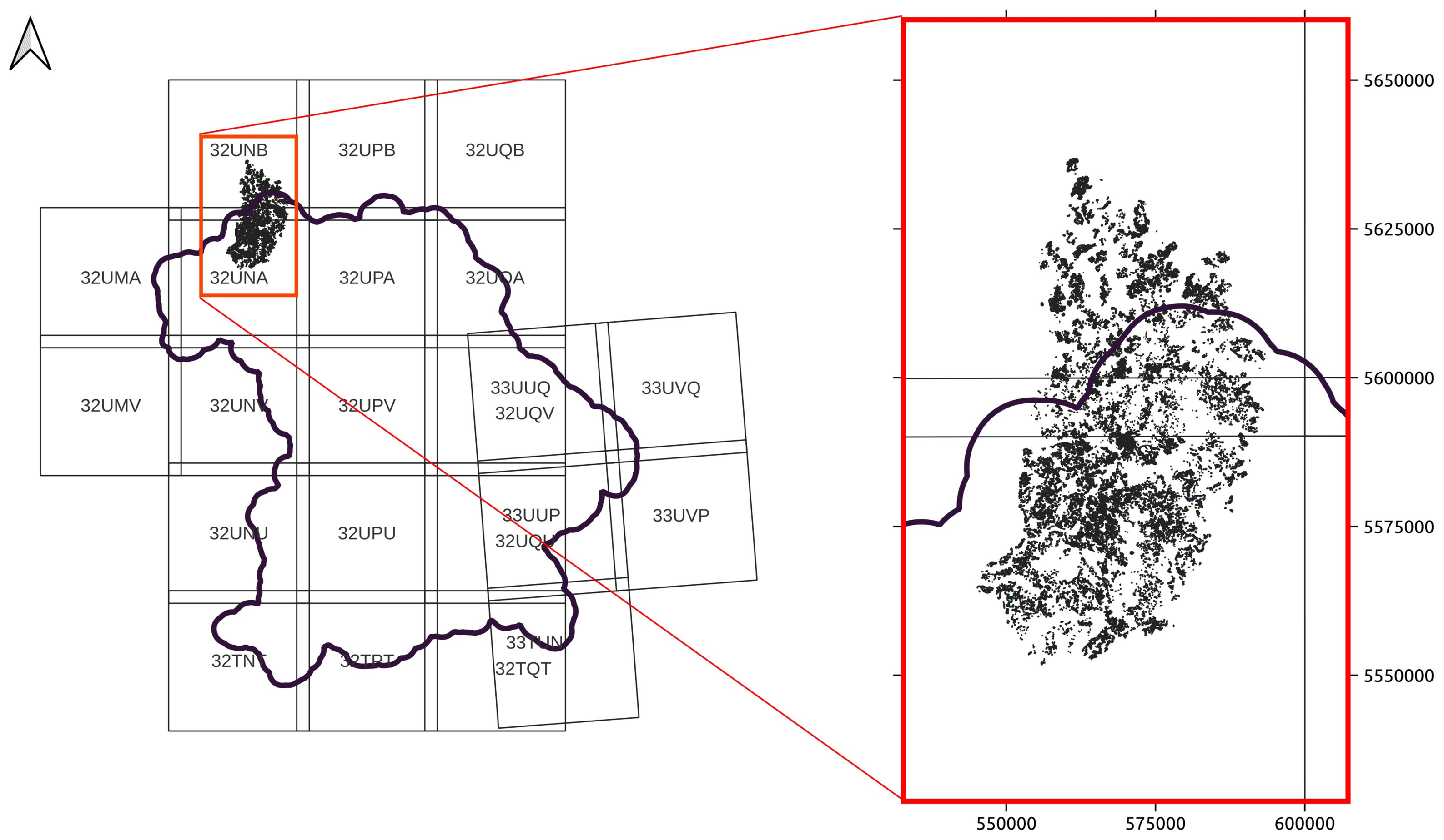

Fig. 1: Sentinel-2 tiles with pure tree stands in the Biosphere Reserve Rhoen (East Hessen, South Thuringia, Northern Bavaria) used as ground truth information.

2. Methods

Based on cloud masked median data for April 2019 we used spectral bands and indices for building a Random Forst model, after combining remote sensing (RS) information with ground truth. April was choosen because of the presence of seasonal signals.
Then we calculated balanced datasets [1] by using two oversampling and two undersampling techniques. Random oversampling (ROS) and a support vector machine based borderline synthetic minority oversampling (SMOTE). Random undersampling (RUS) and condensed nearest neighbor (CNN) undersampling. In addition, the original data set were also used for model building (NO).

Tab. 2: Classes and quantity of the ground truth data from the Rhoen Biosphere Reserve

Tree species	Absolute [N] (relative [%])	Absolute [km^2] (relative [%])
Beech	5667 (30.60)	120.95 (47.14)
Birch	133 (0.72)	0.80 (0.31)
Oak	362 (1.95)	9.74 (3.80)
Alder	149 (0.80)	0.60 (0.23)
Spruce	10209 (55.12)	96.56 (37.63)
Pine	1482 (8.00)	24.97 (9.73)
Larch	519 (2.80)	2.95 (1.15)
Sum	18521 (100.00)	256.57 (100.00)

3. Results

3.1 Model Accuracies

Overall Accurracy (OA) April: 0.79			
	Precision	Recall	F1
Beech	0.76	0.87	0.81
Birch	0	0	0
Oak	0.7	0.08	0.14
Alder	0	0	0
Spruce	0.79	0.88	0.83
Pine	0.71	0.26	0.38
Larch	0	0	0

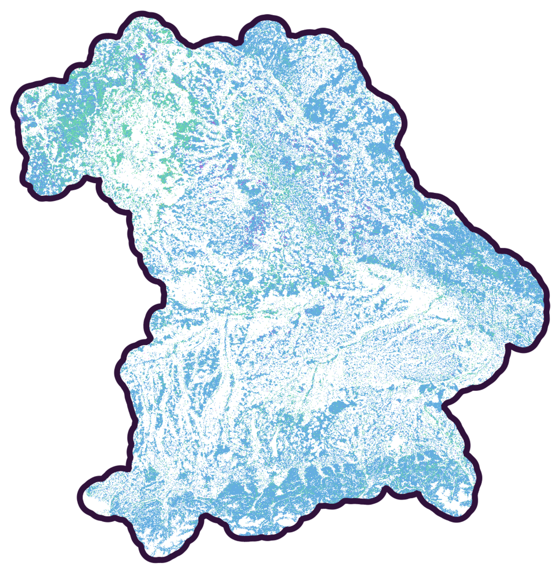

Fig. 3: Model validation metrics of NO data (original ground truth) with resulting map.

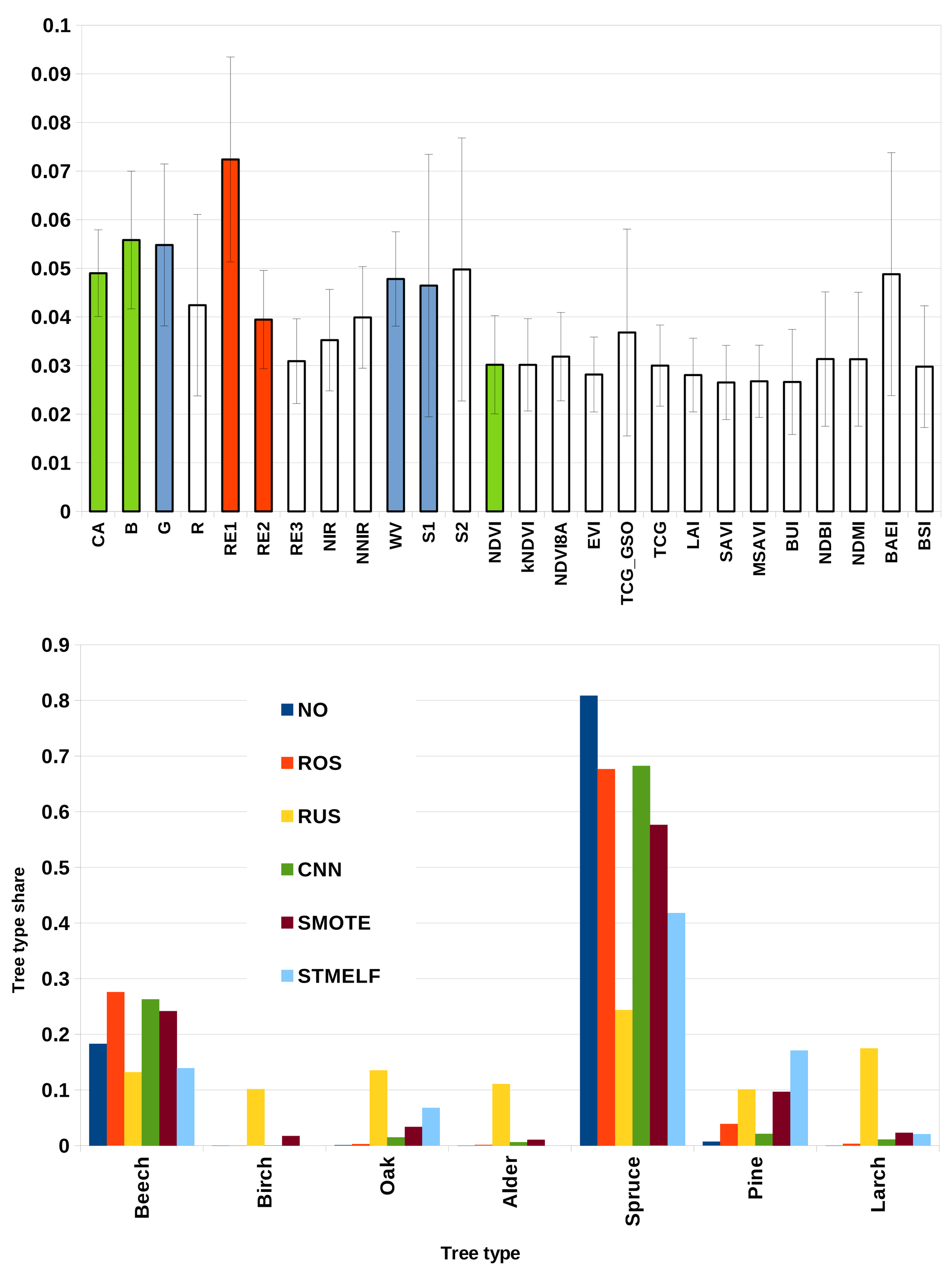

Fig. 4: Top: Gini importance with estimators standard deviation and 5-fold cross validated chosen features by forward (green), backward (blue) and both (red).
Bottom: Comparison with official tree type share [2]

3.2. Classifications

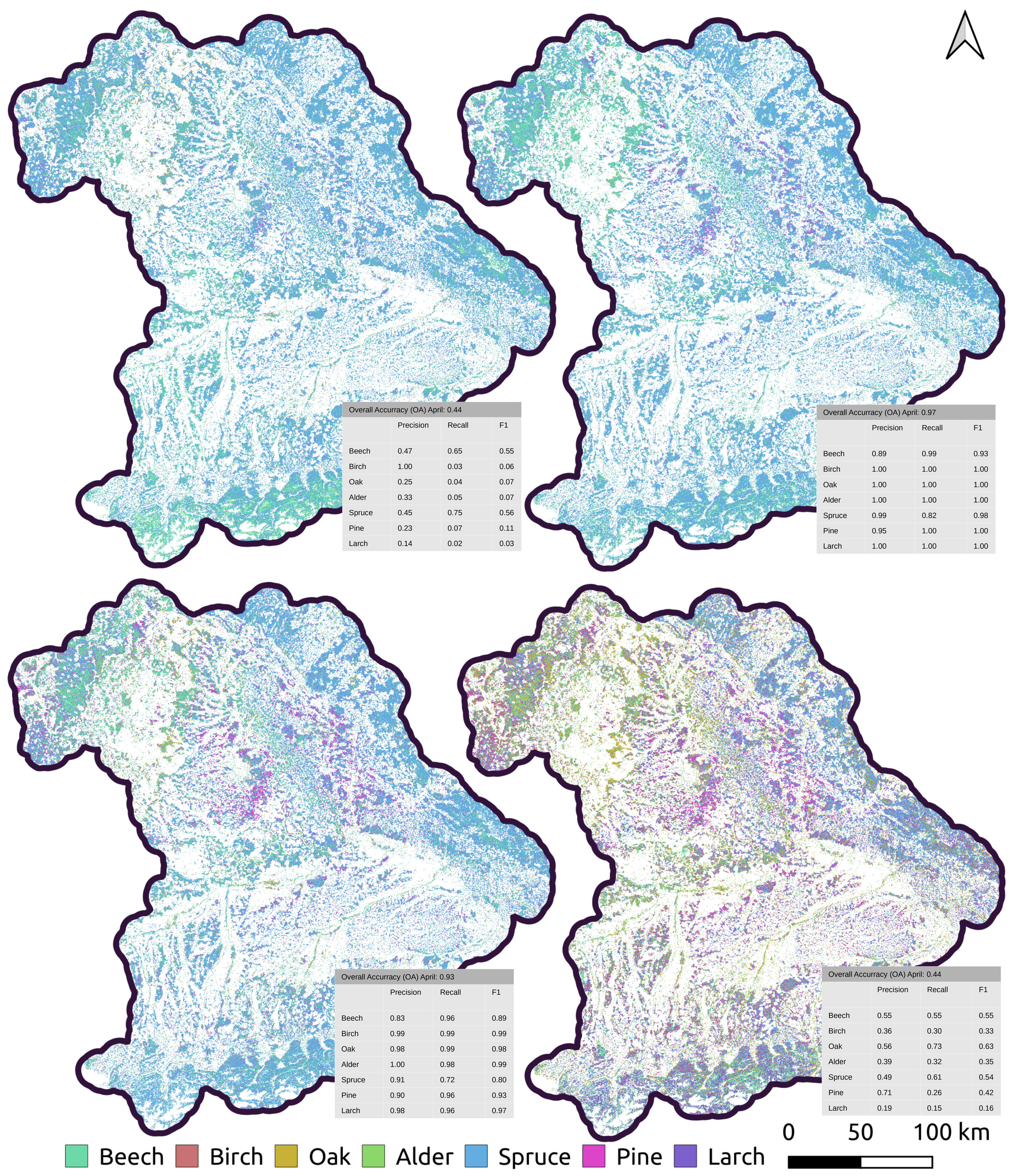

Fig. 5: Resulting maps from different preprocessing steps. Upper left:CNN, upper right: ROS, lower left: SMOTE, lower right: RUS;

4. Conclusion and Oulook

Explaining potential of accurracy metrics is limited concerning final classifications. Even mapping of tree types with with zero accuracy were observed (Fig. 3 birch, alder, larch) and feature importances not totally in line with selection methods. Balancing has a clear impact on the final maps, while overfitting was also observed in the model evaluation. However, when comparing the classification results with official data from other sources [1] the overall tree type share is best explained by using SMOTE data with an overestimation of most frequent ground truth tree types beech and spruce.
Next steps will be performing the classification task with more time information and comparing results with forest inventory data for checking the infuence of single an multi time step satellite data.

Literature:

[1] G., Lemaitre, F., Nogueira and C.K. Aridas. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. Journal of Machine Learning Research. 2017, Vol 18. Nr. 17. Pages 1-5. http://jmlr.org/papers/v18/16-365.html
[2] STMELF. Bayerisches Staatsministerium für Ernährung, Landwirtschaft und Forsten. Wald in Zahlen 2022. URL: https://www.stmelf.bayern.de/wald/forstverwaltung/wald-in-zahlen-fakten-ueber-bayerns-waelder/index.html, accessed 01.06.2023;

The resarch was realized by the funding of the Bavarian State Ministry for Health and Care with the funding number G55b-G8175.0-2017/9-9. Thank you very much!
Special thanks to the administration of the UNESCO biosphere reserve Rhoen for essentially supporting this research by sharing an extensive classification product!