Scaling crop yield estimation: Leveraging earth observation and machine learning ensembles on CODE-DE

Patric Brandt[1], Florian Beyer[1], Peter Borrmann[1,2], Markus Möller[1], Heike Gerighausen[1]

Julius Kühn-Institut
Federal Research Centre for Cultivated Plants

[1] Julius Kühn Institute (JKI) -- Federal Research Centre for Cultivated Plants, Institute for Crop and Soil Science, Bundesallee 58, 38116 Braunschweig, Germany
[2] Current address: Helmholtz Centre for Environmental Research (UFZ), Department of Remote Sensing, Land Cover and Dynamics, Permoserstraße 15, 04318 Leipzig, Germany

Introduction

Estimating crop yields is pivotal for official statistics on agricultural productivity to inform policy-making on sustainable food production. Currently, official crop yield statistics in Germany relies on extensive and time-consuming farm surveys and on-farm measurements.
EU's Copernicus earth observation (EO) program provides a plethora of satellite data, enabling the remotely sensed monitoring of agricultural land at high spatio-temporal resolution. EO imagery, geospatial data on meteorological conditions and soil properties as well as advances in machine learning (ML) provide huge opportunities for model based crop yield estimation, covering large spatial scales with unprecedented granularity.
This study estimates yields multi-annually, covering four major crops in Germany, using ML ensembles and multi-source geodata leveraging the EO cloud platform CODE-DE.

Approach

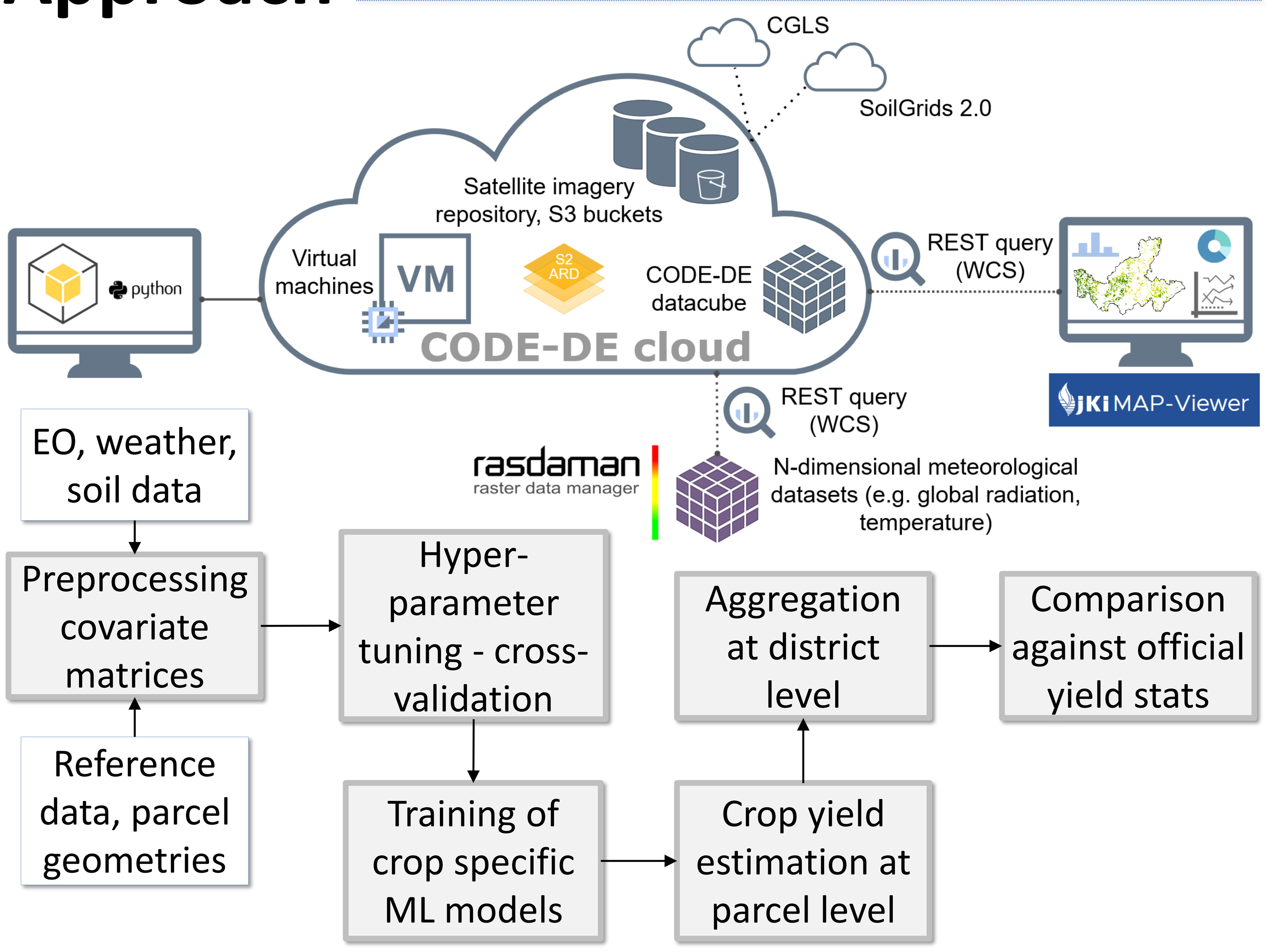


Fig 1. CODE-DE infrastructure (top) and implemented modelling work flow (bottom).

Results

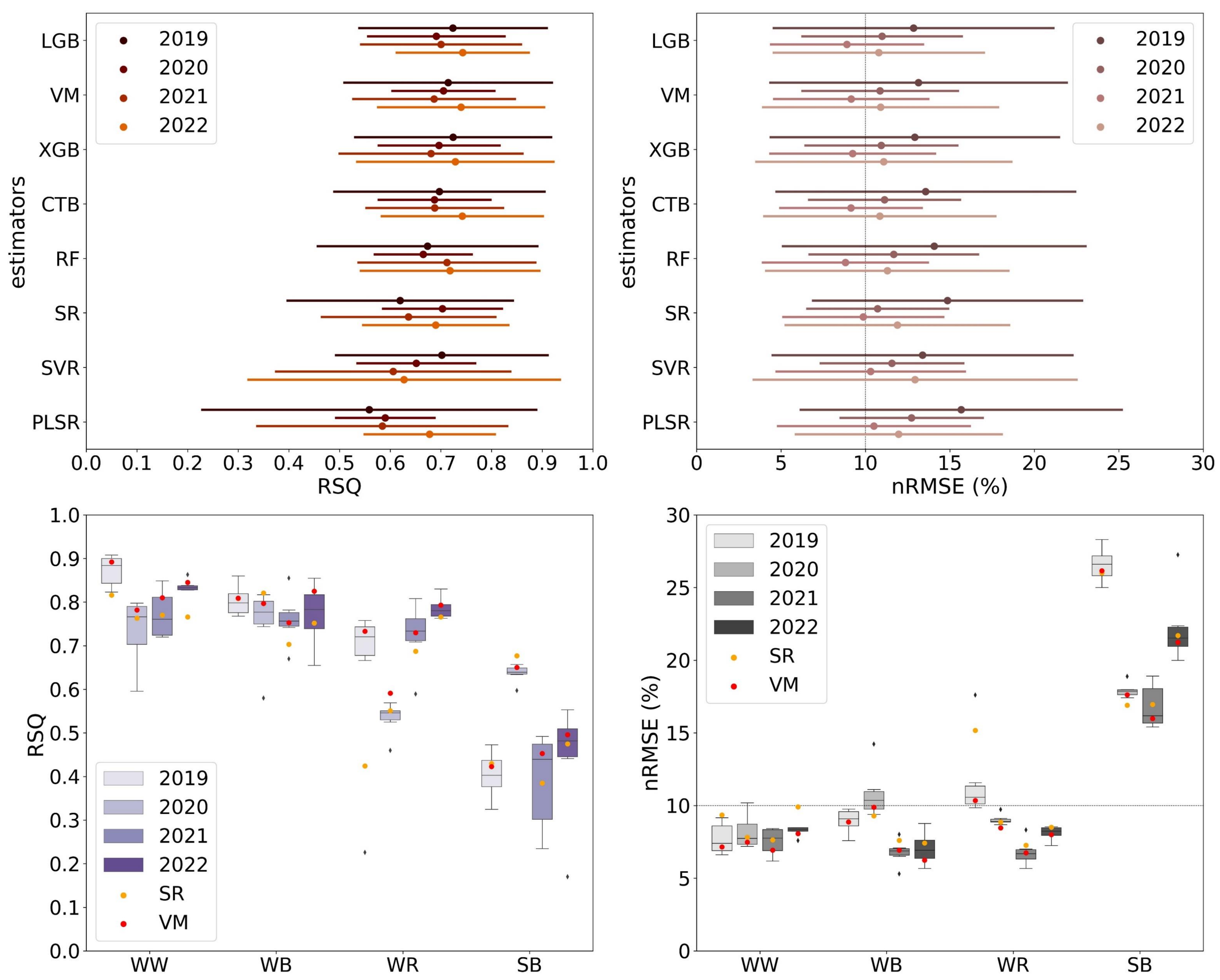


Fig 2. Evaluation of aggregated crop yield estimations over four years for two federal states using an ensemble of ML estimators (upper half per estimator, lower half per crop).

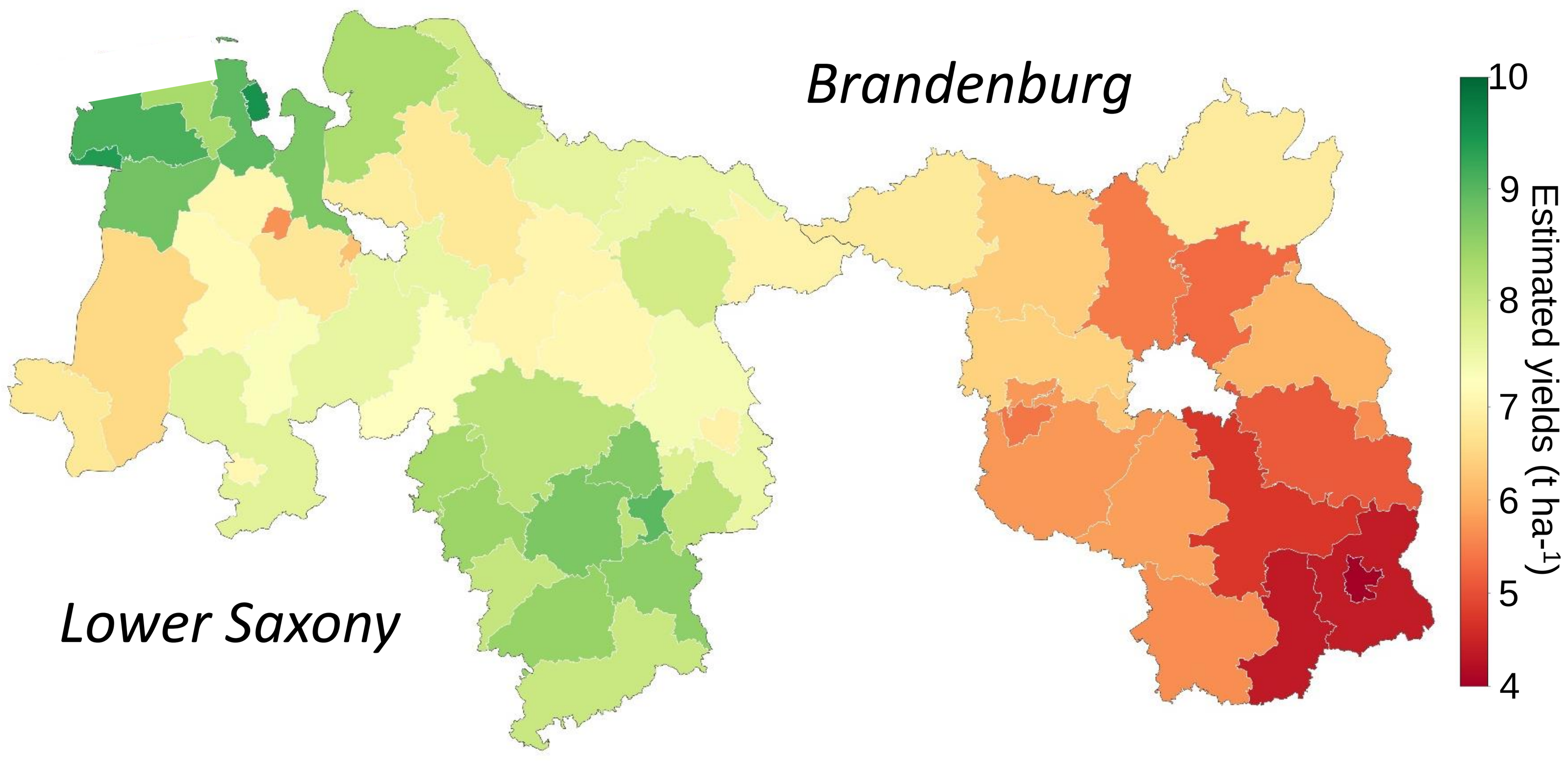


Fig 3. Example map depicts winter wheat yields estimated for 2020, aggregated at district level for two federal states based on approx. 80,000 parcels.

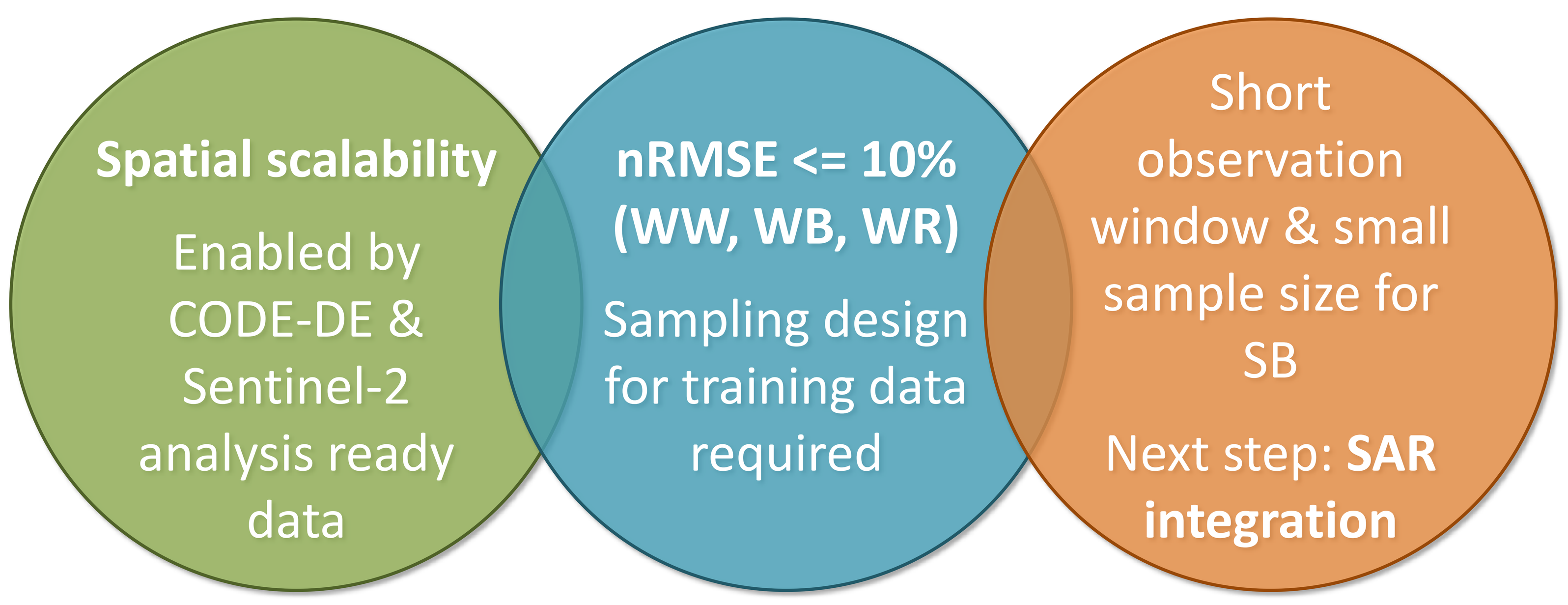
ML ensemble:				Targeted crops:		Evaluation metrics:	
Base estimators		*Meta estimators*					
CTB	CatBoost	SR	ElasticNet	WW	winter wheat	RSQ	Coefficient of determination
LGB	LightGBM			WB	winter barley		
PLSR	Partial Least Square	VM	Voting mean *(histogram-based majority voting)*	WR	winter rape	nRMSE	normalized root mean square error
RFR	RandomForest			SB	spring barley		
SVR	Support Vector						
XGB	XGBoost						

Discussion

- Estimating yields for approx. 160,000 parcels per year spread across two federal states demonstrates capability to scale by using CODE-DE.
- nRMSEs of best models range between 5 – 10% for winter crops and between 15 – 25% for spring barley.
- LightGBM outperformed other ensemble estimators, including the meta estimators.
- Ensemble yield estimations were further used to apply principle of majority voting ascertaining parcel-wise means of 'most trusted' yield estimations (Fig. 2).
- Robust performance of ensemble-based majority voting suggests operational utility for agricultural statistics.

Conclusions

Proof of concept | Comparison against official yield stats | Yield estimation caveat



Project: SatAgrarStat_PLUS | **Project partners**: German Federal Statistical Office (Destatis) and six statistical state offices | **Funding**: Destatis

Research Centre for Agricultural Remote Sensing

www.julius-kuehn.de