MAPPING CANOPY-LEVEL CROP TRAITS USING TOP-OF-ATMOSPHERE SENTINEL-2 DATA IN GOOGLE EARTH ENGINE

J. ESTÉVEZ[1], K. BERGER[1,2], M. SALINERO-DELGADO[1], L. PIPIA[3], J. VICENT[4], J.P. RIVERA-CAICEDO[5], M. WOCHER[2], P. REYES-MUÑOZ[1], G. TAGLIABUE[6], M. BOSCHETTI[6], J. VERRELST[1]

[1]Image Processing Laboratory (IPL), University of Valencia, Paterna 46980, Valencia, Spain.

[2]Department of Geography, Ludwig-Maximilians-Universität München (LMU), 80333 Munich, Germany.

[3]Institut Cartogràfic i Geològic de Catalunya (ICGC), Parc de Montjüic, 08038 Barcelona, Spain.

[4]Magellium, 31520 Toulouse, France.

[5]Secretary of Research and Postgraduate, CONACYT-UAN, Tepic 63155, Mexico.

[6]Institute for Electromagnetic Sensing of the Environment, National Research Council of Italy, 20133 Milano, Italy.

Correspondence to: José Estévez[1] (jose.a.estevez@uv.es)

https://doi.org/10.31490/9788024846026-15

Abstract

To take advantage of the vast amount of remote sensing data, cloud computing platforms such as Google Earth Engine (GEE) open new possibilities to develop crop trait retrieval models applicable to any corner of the world. In the present study, we implemented hybrid models directly in GEE for processing Sentinel-2 (S2) top-of-atmosphere (TOA) reflectance data into crop traits. To achieve this, a training dataset was generated using the leaf-canopy RTM PROSAIL in combination with the atmospheric model 6SV. Gaussian processes regression (GPR) retrieval models were then established for 4 canopy-level crop traits namely: leaf area index, canopy chlorophyll content, canopy water content and canopy dry matter content. Successful reduction of the training dataset by 78% was achieved using the active learning technique Euclidean distance-based diversity (EBD). The EBD-GPR model showed moderate to good performance against in situ data over an independent study site (Grosseto, Italy). Obtained maps compared against ESA Sentinels' Application Platform (SNAP) vegetation estimates showed high consistency of both retrievals. Finally, local and national scale maps were successfully generated in GEE, with additionally providing uncertainties. In summary, the proposed retrieval workflow demonstrates the possibility of routine processing S2 TOA data into crop trait maps at any place on Earth as required for operational agricultural applications.

Keywords: biophysical and biochemical crop traits, top-of-atmosphere reflectance, Sentinel-2, Gaussian processes (GP), atmosphere radiative transfer model, Google Earth Engine, Active Learning (AL), hybrid retrieval methods, Euclidean distance-based diversity (EBD), uncertainty estimates

INTRODUCTION

The unprecedented availability of optical satellite data in cloud-based computing platforms, such as Google Earth Engine (GEE), opens new possibilities to develop crop trait retrieval models from the local to the planetary scale. In particular, the optical data provided by Sentinel-2 (S2) constellation is very convenient for agricultural applications. GEE includes the S2 MSI collection, providing a powerful computational capability for mapping vegetation at any corner of the world. Hybrid retrieval models are of most interest to run on these platforms as they combine the advantages of physically-based radiative transfer models (RTM) with the flexibility of machine learning regression algorithms (MLRAs). Among the MLRAs, Gaussian Processes regression (GPR, Rasmussen and Williams (2006)) models excel with fast and excellent retrieval performance, and in addition they provide associated uncertainties of the mean predictions, to be used as quality indicators.

Despite their diversity, the large majority of developed retrieval methods exploited bottom-of-atmosphere (BOA) reflectance, i.e., after an atmospheric correction algorithm has been applied to acquired top-of-atmosphere (TOA) radiance or reflectance. To avoid the uncertainty that the atmospheric correction step may introduce an alternative approach is to upscale training data simulations from canopy to atmosphere levels and derive the vegetation variables directly from TOA data. TOA retrieval methods usually rely on the coupling of a vegetation RTM with an atmosphere RTM (Estévez et al., 2020, Estévez et al., 2021, Estévez et al., 2022), with the latter explicitly modeling the atmospheric effects on the radiance received by the sensor.

GEE provides powerful computational capability for planetary-scale data processing, however, GPR currently is not part of the GEE environment. Only recently, Pipia et al. (2021) developed a workflow that introduced GPR models in the GEE cloud platform. To make this integration possible it was necessary to review the standard GPR regression formulation to achieve a factorization suitable for parallel computing, and to implement the corresponding matrix algebra transformation. Secondly, active learning (AL) techniques (Verrelst et al. 2016) were introduced to reduce the original training dataset without loss of information.

With this ambition in mind, this study strives for facilitating operational processing of TOA-scale developed GPR retrieval models by enabling implementation in the GEE framework.

METHODOLOGY

Our retrieval strategy firstly consisted of the development of hybrid models to estimate crop traits from S2 TOA data within the in-house developed Automated Radiative Transfer Models Operator (ARTMO) software environment (Verrelst et al. 2012). Second, retrieval models were integrated in GEE.

The workflow with corresponding models is shown schematically in Fig. 1. The main steps are: (1) hybrid model development, (2) optimizing the training dataset, and (3) implementation in GEE. Also, additional steps were necessary to conduct the methodology: ground validation and comparison against SNAP estimates.

For the development of the hybrid models for TOA data we generated a training database with the models PROSPECT-4 (Feret et. al 2008) and 4SAIL (Verhoef and Bach, 2007) (PROSAIL). Subsequently, the atmospheric RTM 6SV (Vermote et al. 1997) was coupled with the PROSAIL simulated vegetation spectra to finally obtain TOA radiance data.

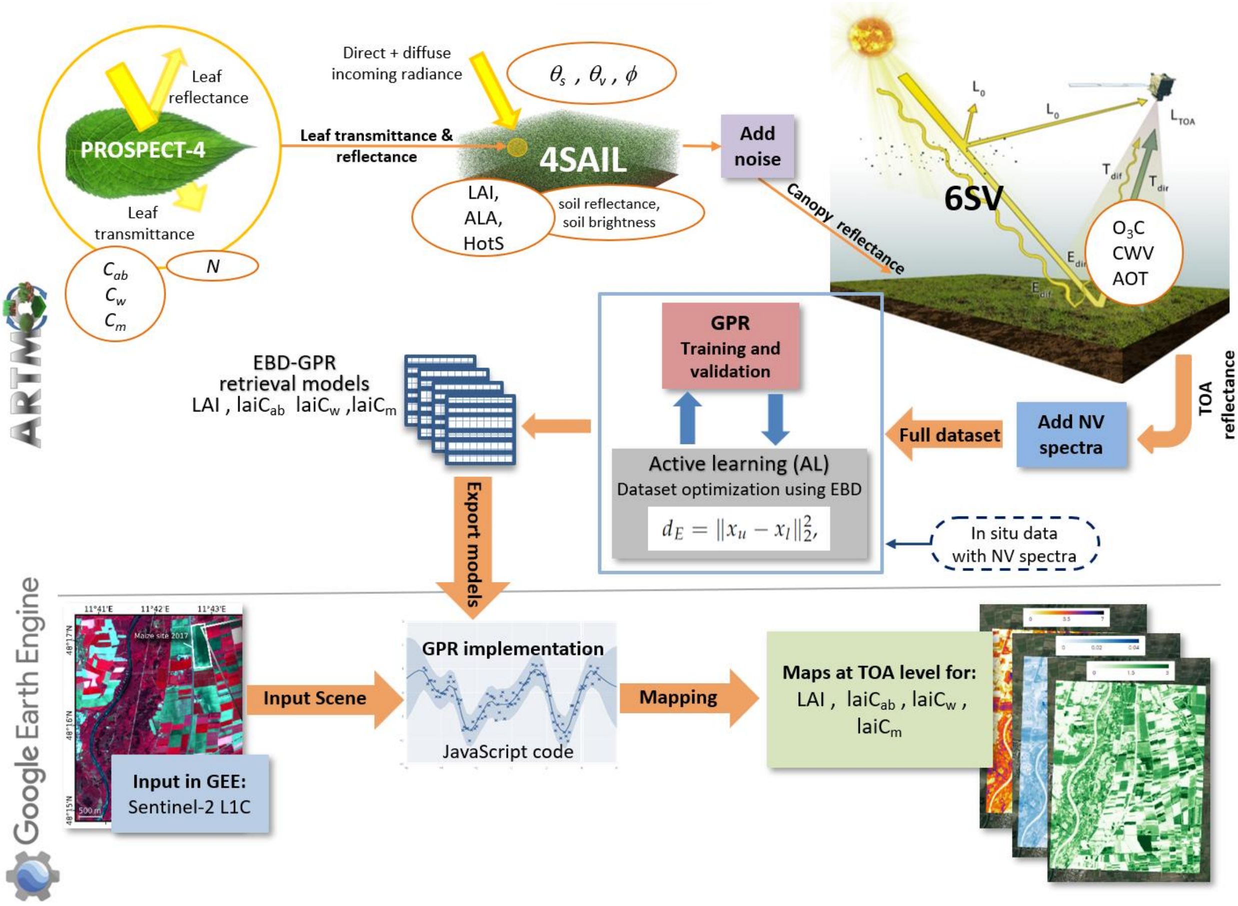

Fig. 1. Flowchart of the pursued workow. Top: Generation of training dataset by coupling of leaf-canopy-atmosphere RTMs for AL optimization and GPR algorithm training. Bottom: Integration of AL-optimized GPR models (EBD-GPR) in GEE for the retrieval of multiple crop traits from S2 TOA data. NV is for non-vegetated.

Furthermore, white Gaussian noise was injected to the simulated PROSAIL spectral data pool, in order to introduce more realism to the RTM data and to prevent overfitting of the models (Brede et al. 2020). Subsequently, 40 spectral samples from non-vegetated surfaces (e.g., water bodies, bare soil or man-made) were added to the training dataset. This step is essential to adapt the models towards processing of full heterogeneous scenes.

Finally, a random dataset of 1'000 simulations of TOA reflectance data was obtained. Hence, the EBD method was applied to optimize the training database, obtaining individual reduced datasets for each variable, i.e., leaf area index (LAI) and upscaled leaf variables canopy chlorophyll content ($laiC_{ab}$), canopy water content ($laiC_w$) and canopy dry matter content ($laiC_m$). These datasets were used for training a GPR generating final retrieval models for each variable were established (EBD-GPR models).

The Munich-North-Isar (MNI) campaigns in Southern Germany (N 48°, E 11°) were explored and directly involved in the AL procedure as a validation dataset. The dataset is composed of structural and biochemical crop variables, which were collected concurrently with field spectroscopic measurements on winter wheat (*Triticum aestivum*) and corn (*Zea maize*) during the 2017 and 2018 growing seasons.

The second dataset for independent validation of the AL-tuned models was collected during an extensive campaign conducted in central Italy (N 42°49.78', E 11°4.21') in the summer of 2018. The site is an agricultural area cultivated with corn and located North of the city of Grosseto.

On the processing side, the Python Application Programming Interface (API) package *ee* provides functions in GEE that allow to extract any available information layer, e.g., S2 Level-1C TOA reflectance (S2-L1C), over a specific area of interest (AOI) and process the resulting datasets very efficiently.

The maps of the different functional vegetation traits were obtained by importing the corresponding EBD-GPR models generated in ARTMO in the GEE environment and performing the mean value prediction on-the-fly. As a result, the GPR mean value retrievals from a specific S2 tile can be visualized in a few seconds at any zoom level, and the maximum resolution map can be downloaded locally in a few minutes. The GEE codes to run the developed models and map the variables can be found in https://github.com/esjoal/GEE_GPR_mapping_vegetation

RESULTS

Successful reduction of the training dataset by 78% was achieved using the active learning technique Euclidean distance-based diversity (EBD). With the EBD-GPR model, highly accurate validation results of LAI and upscaled leaf variables were obtained against *in situ* field data from the validation study site Munich-North-Isar (MNI), with normalized root mean square errors (NRMSE) from 6% to 13%.

Fig. 2 provides the independent validation results of the established retrieval models using the corn dataset from the Grosseto site, Italy. Overall retrieval accuracy can be seen as moderate with NRMSE > 20% (for laiC$_w$ and laiC$_{ab}$) to good with NRMSE < 20% (LAI and laiC$_m$).

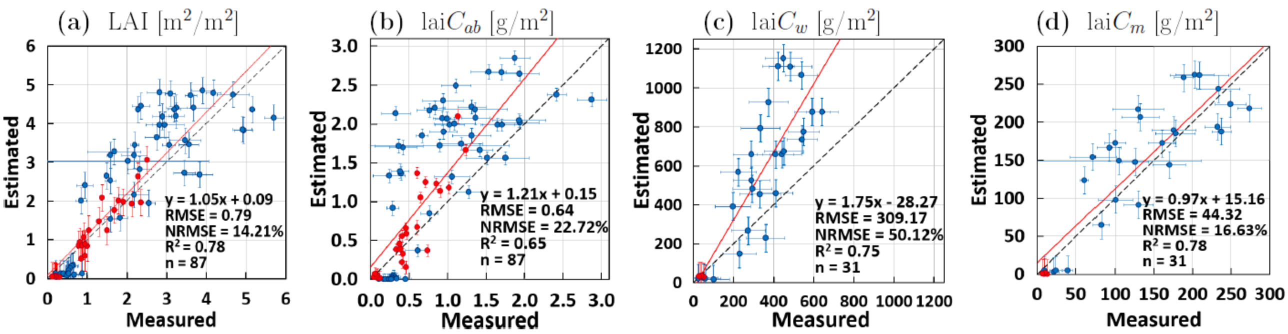

Fig. 2. Ground validation of corn for retrieval of canopy-level crop traits over Grosseto site by the EBD-GPR models from S2-L1C (TOA) reflectance: LAI (a), laiC$_{ab}$ (b), laiC$_w$ (c) and laiC$_m$ (d). Measured vs. estimated values along the 1:1-line. Horizontal bars indicate SD for ground measurements. Vertical bars indicate associated uncertainty estimates (1 SD) for EBD-GPR model. A trend line (red) was added to represent the pattern of the points.

The maps were generated in GEE applying the EBD-GPR models for TOA data products over a subset from the MNI test site on 6^{th} July 2017, as shown in Fig. 3.

The canopy-level models were compared against the same retrievals obtained by the SNAP Sentinel-2 Level 2 Prototype Processor (SL2P), providing high consistency of both retrievals (Fig. 4).

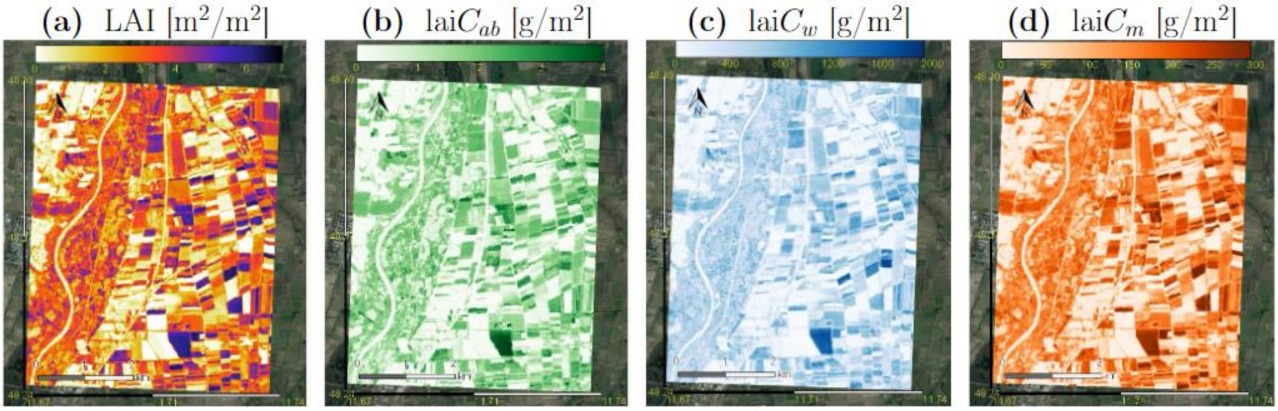

Fig. 3. Maps (mean estimates; μ) of several crop traits: LAI (a), laiC_{ab} (b), laiC_w (c) and laiC_m (d), as generated by EBD-GPR models applied in GEE from S2-L1C data at MNI test site on 6th July 2017.

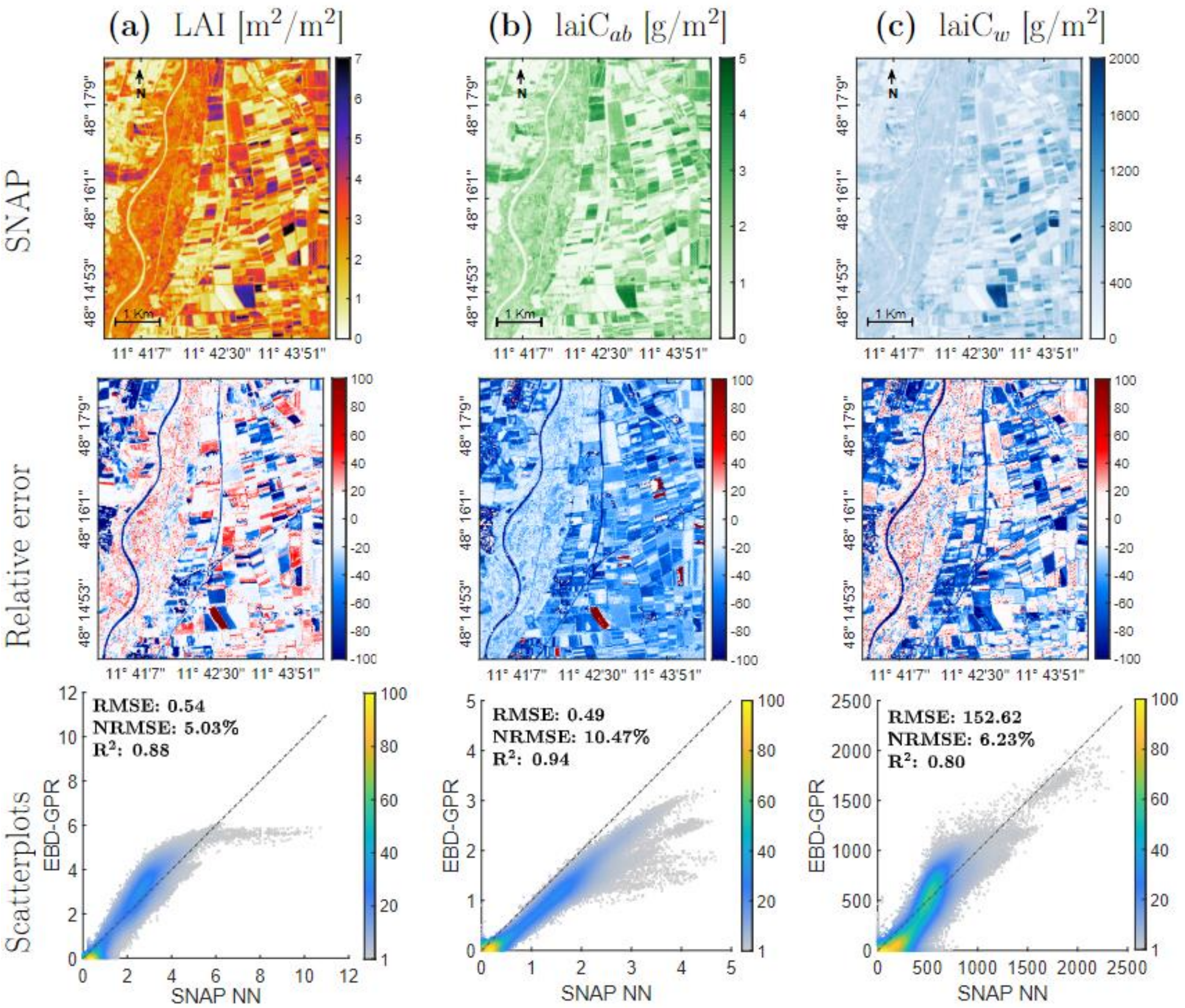

Fig. 4. Crop traits maps as obtained by SNAP SL2P NN (Top), relative error maps (Center) and density scatter plots (Bottom) between EBD-GPR (Fig. 3) and SL2P NN, estimated from S2 Level-2A (BOA) data for LAI (a), laiC_{ab} (b) and laiC_w (c) over MNI site on 6th July 2017. Relative errors and density in %.

To demonstrate large scale mapping capabilities of GEE, we simulated the crop traits at the national scale for the entirety of Germany (Fig. 5). In this case, we considered a time span instead of a specific date, and finally applied the statistical median estimator to obtain a spatially continuous coverage. At a glance, the maps show spatial patterns across the country surface and estimation range of variables seems to be correct.

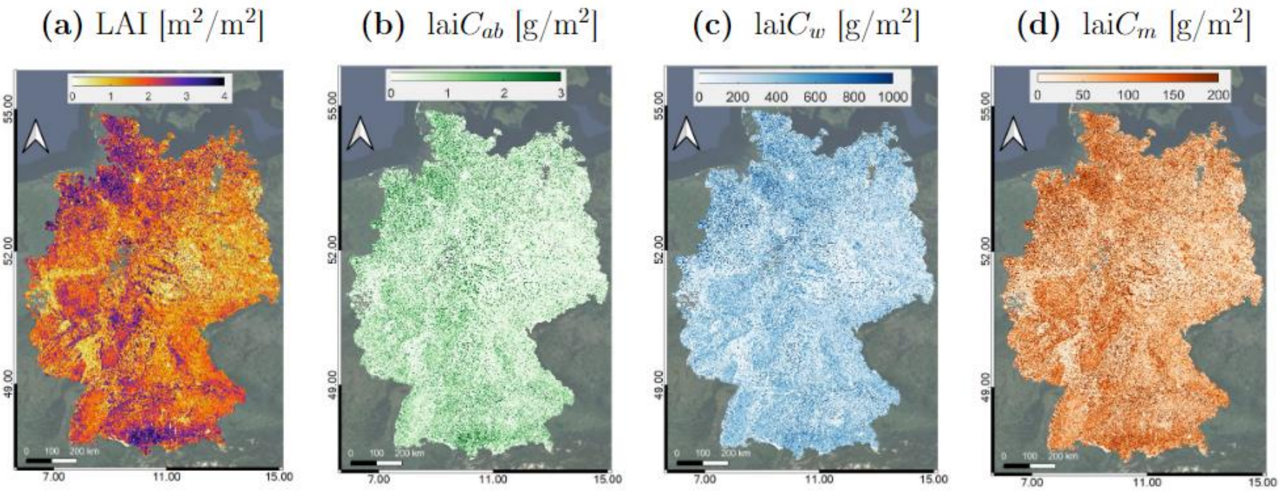

Fig. 5. Maps (mean estimates; μ) at big scale of several crop traits over the whole of Germany: LAI (a), $laiC_{ab}$ (b), $laiC_w$ (c) and $laiC_m$ (d) , as generated by EBD-GPR models applied in GEE from S2-L1C data. Time span covers from 1st to 31st July 2017 using median value strategy.

Finally, thanks to the seamless GEE processing capability, the TOA-based mapping was applied over the entirety of Germany at 20m spatial resolution including information about prediction uncertainty (Fig. 6). The obtained maps provide confidence in the developed EBD-GPR retrieval models for integration into the GEE framework and national scale mapping from S2 L1C imagery.

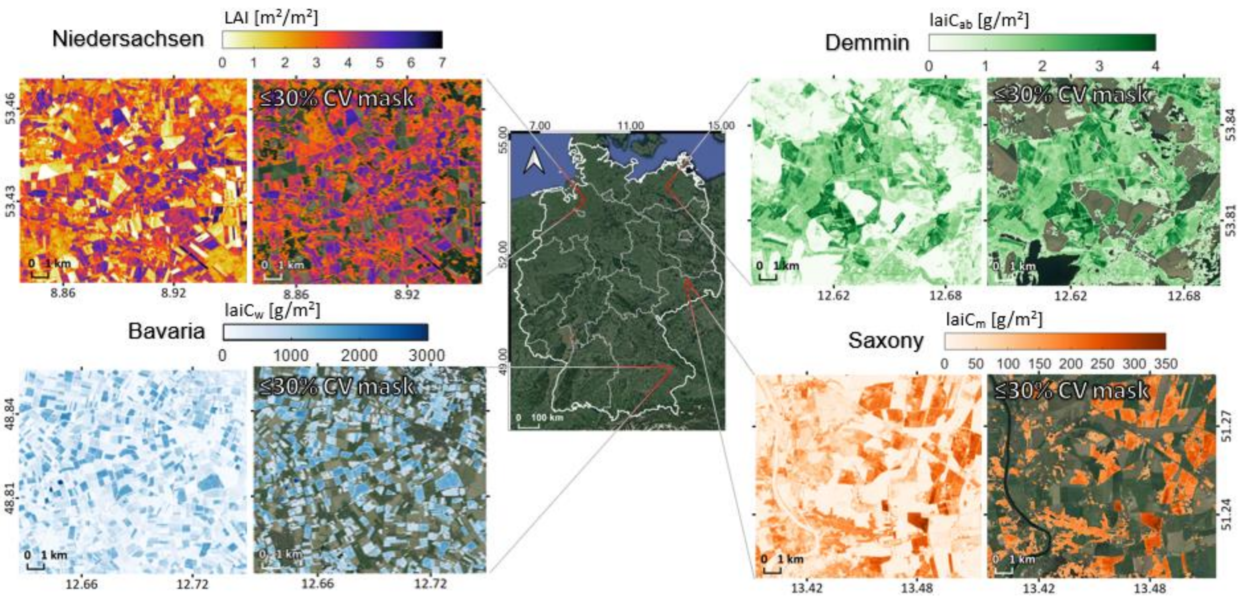

Fig. 6. Subset maps for four selected German regions with typical agricultural usage in 20 m spatial resolution (left maps). Uncertainty provided by the EBD-GPR models was used to mask out areas with more than 30 % of relative uncertainty (CV) (right maps).

CONCLUSIONS

We can conclude that the workflow described here presents a promising path towards operational mapping of essential crop traits with the high-performance computing capacity of GEE and Sentinel-2 data. The proposed approach has the potential to be used for a multitude of agricultural applications supporting management decisions from farm to regional levels.

REFERENCES

Brede, B., Verrelst, J., Gastellu-Etchegorry, J.P., Clevers, J.G.P.W., Goudzwaard, L., den Ouden, J., Verbesselt, J., Herold, M., (2020) Assessment of Workflow Feature Selection on Forest LAI Prediction with Sentinel-2A MSI, Landsat 7 ETM+ and Landsat 8 OLI. Remote Sens. 12.

Estévez, J., Vicent, J., Rivera-Caicedo, J.P., Morcillo-Pallarés, P., Vuolo, F., Sabater, N., Camps-Valls, G., Moreno, J., Verrelst, J., (2020) Gaussian processes retrieval of LAI from Sentinel-2 top-of-atmosphere radiance data. ISPRS J. Photogramm. Remote Sens. 167, 289-304.

Estévez, J., Berger, K., Vicent, J., Rivera-Caicedo, J.P., Wocher, M., Verrelst, J., (2021) Top-of-Atmosphere Retrieval of Multiple Crop Traits Using Variational Heteroscedastic Gaussian Processes within a Hybrid Workflow. Remote Sens. 13(8), 1589.

Estévez, J., Salinero-Delgado, M., Berger, K., Pipia, L., Rivera-Caicedo, J.P., Wocher, M., Reyes-Muñoz, P., Tagliabue, G., Boschetti, M., Verrelst, J., (2022) Gaussian processes retrieval of crop traits in Google Earth Engine based on Sentinel-2 top-of-atmosphere data. Remote Sens. Environ. 273, 112958

Feret, J.B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S.L., le Maire, G., Jacquemoud, S., (2008) PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. Remote Sens. Environ. 112, 3030-3043

Pipia, L., Amin, E., Belda, S., Salinero-Delgado, M., Verrelst, J., (2021) Green LAI Mapping and Cloud Gap-Filling Using Gaussian Process Regression in Google Earth Engine. Remote Sens. 13, 403.

Verhoef, W., Bach, H., (2007) Coupled soil-leaf-canopy and atmosphere radiative transfer modelling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data. Remote Sens. Environ. 109, 166-182.

Vermote, E., Tanré, D., Deuzé, J., Herman, M., Morcrette, J.J., (1997) Second simulation of the satellite signal in the solar spectrum, 6S: an overview. IEEE Trans. Geosci. Remote Sens. 35, 675-686.

Verrelst, J., Romijn, E., Kooistra, L., (2012) Mapping vegetation density in a heterogeneous river Floodplain ecosystem using pointable CHRIS/PROBA data. Remote Sens. 4, 2866-2889

Verrelst, J., Dethier, S., Rivera, J.P., Munoz-Mari, J., Camps-Valls, G., Moreno, J., (2016) Active learning methods for e-cient hybrid biophysical variable retrieval. IEEE Geosci. Remote Sens. Lett. 13, 1012-1016.

Verrelst, J., Vicent, J., Rivera-Caicedo, J.P., Lumbierres, M., Morcillo-Pallarés, P., Moreno, J., (2019) Global Sensitivity Analysis of Leaf-Canopy-Atmosphere RTMs: Implications for Biophysical Variables Retrieval from Top-of-Atmosphere Radiance Data. Remote Sens. 11(16), 1923.

Rasmussen, C.E., Williams, C.K.I., (2006) Gaussian Processes for Machine Learning. The MIT Press, New York.