ISPRS Open Journal of Photogrammetry and Remote Sensing 8 (2023) 100033

Contents lists available at ScienceDirect

ISPRS Open Journal of Photogrammetry and Remote Sensing

journal homepage: www.journals.elsevier.com/isprs-open-journal-of-photogrammetry-and-remote-sensing

Pixel-based mapping of open field and protected agriculture using constrained Sentinel-2 data

Daniele la Cecilia [a,*,1], Manu Tom [b,c], Christian Stamm [a], Daniel Odermatt [b,c]

[a] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Environmental Chemistry, Überlandstrasse 133, 8600, Dübendorf, Switzerland

[b] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Surface Waters - Research and Management, Überlandstrasse 133, 8600, Dübendorf, Switzerland

[c] University of Zurich, Department of Geography, Winterthurerstrasse 190, 8057, Zurich, Switzerland

ARTICLE INFO

Keywords:
Protected agriculture
Plastic greenhouse
Random forest
Sentinel-2
Corine

ABSTRACT

Protected agriculture boosts the production of vegetables, berries and fruits, and it plays a pivotal role in guaranteeing food security globally in the face of climate change. Remote sensing is proven to be useful for identifying the presence of (low-tech) plastic greenhouses and plastic mulches. However, the classification accuracy notoriously decreases in the presence of small-scale farming, heterogeneous land cover and unaccounted seasonal management of protected agriculture. Here, we present the random forest-based pixel-level Open field and Protected Agriculture land cover Classifier (OPAC) developed using Sentinel-2 L2A data. OPAC is trained using tiles from Switzerland over 2 years and the Almeria region in Spain over 1 acquisition day. OPAC classifies eight land covers typical of open field and protected agriculture (plastic mulches, low-tech greenhouses and for the first time high-tech greenhouses). Finally, we assess (1) how the land covers in OPAC are labelled in the Sentinel-2 Scene Classification Layer (SCL) and (2) the correspondence between pixels classified as protected agriculture by OPAC and by the best performing Advanced Plastic Greenhouse Index (APGI). To reduce anthropogenic land covers, we constrain the classification task to agricultural areas retrieved from cadastral data or the Corine Land Cover map. The 5-fold cross-validation reveals an overall accuracy of 92% but other classification scores are moderate when keeping the separation among the three classes of protected agriculture. However, all scores substantially improve upon grouping the three classes into one (with an Intersection Over Union of 0.58 as an average among the scores of the three classes and of 0.98 for one single class). Given the recently acknowledged importance of Sentinel-2 Band 1 (central wavelength of 443 nm), the classification accuracy of OPAC for the Swiss small-scale farming is mostly limited by the band's reduced spatial accuracy (60 m). A careful visual assessment indicates that OPAC achieves satisfactory generalization capabilities also in North European (the Netherlands) and four Mediterranean areas (Spain, Italy, Crete and Turkey) without the need of adding location and temporal specific information. There is good agreement among natural land covers classified by OPAC and the SCL. However, the SCL does not have a class for protected agriculture, the latter being often classified as clouds. APGI achieved similar to lower classification accuracies than OPAC. Importantly, the APGI classification task depends on a user-defined space- and time-specific threshold, whereas OPAC does not. Therefore, OPAC paves the way for rapid mapping of protected agriculture at continental scale.

1. Introduction

Protected agriculture is a promising technological solution to securely produce vegetables, berries and fruits for a growing human population in the face of varying meteorological conditions driven by climate change. Protected agriculture comprises low- and high-tech crop covers that modify local environmental factors (rainfall, irrigation, temperature, radiation, wind, humidity, and thus evapotranspiration) to secure larger yields. Because protected agriculture is highly remunerative and supports the economic development of rural areas, it has been spreading globally at an increasing rate (Briassoulis et al., 2016; Espí et al., 2006).

* Corresponding author. Eawag Department of Environmental Chemistry, Überlandstrasse 133, 8600, Dübendorf, Switzerland.
E-mail address: daniele.lacecilia@eawag.ch (D. la Cecilia).
[1] Now at: Eawag, Swiss Federal Institute of Aquatic Science and Technology, Water Resources and Drinking Water, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland.

https://doi.org/10.1016/j.ophoto.2023.100033
Received 2 December 2022; Received in revised form 9 February 2023; Accepted 23 February 2023
Available online 27 February 2023

Low-tech plastic sheets for covering soils (mulches) and/or plants (tunnels and greenhouses) are the most common. They are used either seasonally or year-round and they serve to reduce evaporation, increase soil temperatures or avoid germination of unwanted weeds. Direct environmental impacts of the use of plastic sheets include the large amount of plastic waste and soil pollution from microplastics and their chemical additives (Harms et al., 2021; Huang et al., 2020; Lian et al., 2022; Wang et al., 2021). On the other hand, high-tech protected agriculture implies the construction of built structures equipped with systems that regulate environmental factors. Here, crops are often grown out of the soil and directly within aqueous growing media that provide the right dose of micro- and macro-nutrients. Such greenhouses are most beneficial where natural conditions are not suitable for growing vegetables and berries, such as in light-scarce and extreme climates. High-tech greenhouses may have high energy demand to provide light and regulate large indoor volumes (Ntinas et al., 2017).

Remote sensing products play a pivotal role in global near real-time agricultural land cover classification. They provide precious information to manage agricultural areas for food security and to drive policy-making for environmental sustainability (Defourny et al., 2019). In the context of plastic agriculture mapping, a plethora of methodologies have been published (M. Á. Aguilar et al., 2020). These methodologies employ indices or classification models to define whether the pixels or objects of the sensed land covers belong to plastic agriculture or not. Indices are calculated as the algebraic combination of selected input data (typically, the reflectance of important wavelengths). The numerical output is then compared against a defined threshold for land cover classification. Differently, classification models are trained on a dataset of numerical features determined to be important for classifying the target land covers. The trained model is then applied to classify new datasets composed of the same features used in the training phase. The most accurate and robust index for mapping plastic greenhouses and mulches against other land covers, based on the exceedance of a threshold value, at the pixel level, is the recent “advanced plastic greenhouse index” (APGI) (Zhang et al., 2022). The APGI is developed for Sentinel-2 data and it achieves overall accuracies ranging from 90% to 97%. However, the authors noted that the threshold for mapping using the APGI is subject to spatial and seasonal variations. Classification models based on machine learning or deep learning algorithms are an alternative methodology to map plastic agriculture (see selected examples in Table 1).

As far as mapping protected agriculture using radar data is concerned, Hasituya et al. (2017) note that polarimetric decomposition features are more important than the backscatter itself. However, Lanz et al. (2020) state that plastic is almost transparent in Sentinel-1’s C band. Several factors might affect the backscatter signal from plastic greenhouses, including the metal frame supporting the plastic sheets and the presence of condensing moisture on the plastic sheets. These factors require more thorough investigations.

Operational satellite-based agriculture monitoring at global scale is advancing quickly (Defourny et al., 2019), but protected agriculture mapping is not accounted for in such applications, as in Tsendbazar et al. (2021). Challenges for global monitoring of protected agriculture arise from spectral ambiguity, spatial resolution and the lack of robust phenological patterns. A first challenge for mapping protected agriculture is determined by spectral ambiguity due to clouds and other highly reflective surfaces. Clouds hinder the use of optical data for any land cover classification application. The wide range of plastic materials used in protected agriculture (Mormile et al., 2017) leads to very high intra-class variations of spectral signatures (Zhang et al., 2022). In contrast, other highly-reflective natural and synthetic surfaces determine very low inter-class variations of spectral signatures. To improve the classification accuracy of plastic greenhouses mapping, there has been an increasing reliance on image segmentation prior to the use of classification models (e.g., references in Table 1). The segmentation step unifies contiguous spatial units based on their similarity in terms of

Table 1
References to different methods and the achieved overall accuracy based on classification models to map plastic agriculture distinguished by the sensor(s) and approach used. When one author study more sites, we calculate the mean overall accuracy.

Author, year	Method	Approach	Sensor(s)	Overall Accuracy[a]
Acharki and Veettil (2022)	Random Forest	Pixel	Landsat-8 and Sentinel-2, PlanetScope and RapidEye	>87%
Zhang et al. (2022)	MDI by Aguilar et al. (2016)	Pixel	Sentinel-2	>89%
	GDI by González-Yebra et al. (2018)			>77%
	PGI by Yang et al. (2017)			>86%
	APGI by Zhang et al. (2022)			>95%
Balcik et al. (2020)	Random Forest	Object	Sentinel-2	>87%
	Support-Vector Machine			>85%
	K-nearest neighbour			>88%
	Random Forest		SPOT-7	>91%
	Support-Vector Machine			≈88%
	K-nearest neighbour			>91%
Nemmaoui et al. (2018)	Decision tree on time series	Object	Landsat-8 and Sentinel-2A	≈93%
Ji et al. (2019)	Threshold model	Object	Landsat-8	>98%
Novelli et al. (2016)	Random Forest	Object	Sentinel-2, Landsat-8 and World-View 2	≈94%
Lu et al. (2018)	Classification and regression tree	Object	Sentinel-2 and Sentinel-1	≈90%
	Random Forest			≈94%
	Support-Vector Machine			≈94%
Sun et al. (2021)	1-D Convolutional Neural Network	Object	Sentinel-2	≈90%
Ma et al. (2021)	Dense object dual-task deep learning	Object	Several sources of RGB images at 1 m resolution	Not calculated

[a] The overall accuracy was often the only metric measured the studies. However, we argue that overall accuracy is a biased metric in the presence of unbalanced training data, that is when different classes contain a different number of training data. (Ma et al., 2021) reported other performance scores.

spectral information and texture. These units allow for adding contextual information, such as shape, which could help to improve the accuracy of the classification model. However, segmentation requires additional expertise and high computational power. Yet, to be highly effective, segmentation requires very high-resolution images in order to allow for distinguishing small objects, thus reducing the effect of mixed land cover within a pixel. However, these images are often not freely available. The need for expertise to perform segmentation and the cost of very high-resolution images hinder the up-scaling of object-based approaches. The second challenge is represented exactly by the presence of small-scale farming because it causes high inter- and intra-plot variability (Bégué et al., 2020). This issue is exacerbated in protected agriculture because plastic greenhouses may occupy a small portion of land or may not be contiguous to let room for machines. Yet, plastic mulches can be spaced out with bare soil or vegetation. Indeed, Hasituya et al. (2017) recommend to use a spatial resolution smaller than 10 m for applications related to plastic agriculture. However, there are no operational satellite sensors achieving such resolution at the wavelengths of Sentinel-2. Rather, Zhang et al. (2022) highlighted the importance of Band 1 of Sentinel-2 for plastic agriculture mapping, which has a

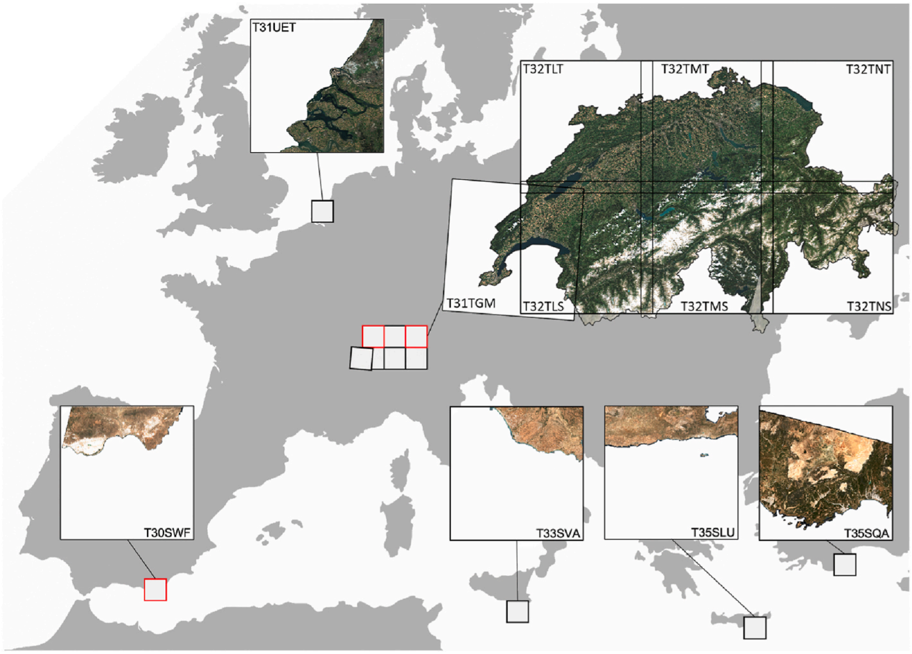

Fig. 1. Study areas. White-filled squares indicate the 100 km × 100 km UTM grid squares in which Sentinel-2 products are subset for distribution. Red-contoured tiles are used for model training and application. Black-contoured tiles are used for model application. Zoomed in tiles show the Sentinel-2 True Color Image at 10 m resolution in 2021 for Switzerland on August 12th (T32TLT, T32TMT, T31TGM, T32TLS, T32TMS and T32TMS) and on August 14th (T32TNT), for the Netherlands on September 7th (T31UET), for Spain on August 28th (T30SWF), for Italy on July 16th (T33SVA), for Greece on July 31st (T35SLU) and for Turkey on August 27th (T35SQA). The grey background is the blank map of Europe without borders available at https://commons.wikimedia.org/wiki/File:Blank_map_europe_no_borders.svg.

resolution of 60 m only. Moreover, mixed land cover within a pixel can be critical in countries with large spatial heterogeneity due to fragmented land covers (i.e., in small-scale farming) or topographical conditions, such as narrow mountain valleys. The third challenge is crop management, which can lead to multiple and faster land cover changes over the agronomical year. These changes could confound the use of time-series analysis for land cover classification (Bégué et al., 2020). Management practices are particularly relevant in protected agriculture where plastic mulching is typically used upon seeding, but in a short period, the growing plants influence the spectral signature of the mulch (Hasituya et al., 2017). As far as plastic greenhouses are concerned, they may be temporarily installed during the growing season to protect fruits from adverse meteorological events and they may also be white-painted to enhance the protection offered to the growing crops by the plastic tunnels against hot temperatures and intense sunlight (M. Aguilar et al., 2014). Finally, all studies have focused on the mapping of low-tech rather than high-tech greenhouses. This latter category of protected agriculture is expanding globally in countries where climatic conditions are adverse for local fruit and vegetable production, which are nutritious in humans' diet (van Delden et al., 2021).

In this research, we perform per-pixel classification of Sentinel-2 L2A products with a random forest algorithm due to its speed, robustness and the possibility to retrieve the reason(s) for a particular land cover assignation. Our choice relies on the concept that the model has learnt the signature of the eight targeted land covers in space and time. Thus, we present the Open field and Protected Agriculture land cover Classifier OPAC for Sentinel-2 data. We train OPAC on 5151 pixels labelled with eight different land covers covering a wide variety of agricultural landscapes, two years of monthly variations in environmental conditions and several different types of protected agriculture covers (i.e., mulches, low-tech greenhouses and for the first time high-tech greenhouses). In addition to 5049 pixels in Switzerland, 102 pixels are sampled and labelled from four relevant land cover classes within the Mediterranean region of Almeria in southern Spain. This aims to adapt to locally specific spectral signatures for the target land covers in other climates, as well as to local cropping management practices (i.e., white-painting). By applying OPAC to single pilot Sentinel-2 products relative to the Netherlands, Spain, Italy, Crete and Turkey, we test whether OPAC can effectively supersede the step of fine-tuning space- and time-specific thresholds needed in the index-based studies for mapping plastic greenhouses using the APGI developed by Zhang et al. (2022) as an example. Moreover, we evaluate whether OPAC can also classify natural land covers through the comparison with the Sentinel-2 Scene Classification Layer (SCL). Finally, instead of image segmentation, we propose to resolve spectral ambiguity using national cadastral inventories or the Corine Land Cover map to constrain the classification effort.

2. Study area and data

2.1. Study areas

We research on the presence of protected agriculture in Switzerland and five other target locations relevant for generalizing our work over Europe (i.e., The Netherlands, Spain, Turkey, Italy and Greece) (Fig. 1).

Switzerland has a complex topography putting forth a wide diversity of agricultural landscapes, which cover about 40% of the territory and amount to 1,044,000 ha. In 2020, agricultural areas were covered by pasture and grasslands for 69.9%, cereals for 13.6%, open field agriculture for 6.9%, oil seeds and fruits for 3.0%, potatoes and beetroots for 2.8%, perennial cultures for 2.3% and others for 1.5% (UST, 2021). Importantly, 1.2% of agricultural areas were destined to horticulture, of which 4% was covered with greenhouses (UST, 2021).

Due to climatic reasons, the Netherlands have focused on vegetable production in high-tech greenhouses. Most of the greenhouses are located in the Hague and in the Rotterdam areas, where cucumbers, tomatoes and sweet peppers are produced (van der Velden et al., 2004). The Sentinel-2 tile for this site is T31UET.

Mediterranean areas are the ones where low-tech plastic-made greenhouses are predominantly found. Worldwide, the most investigated area is the Almería region in Spain. Here, the most economically relevant crop and berry grown below greenhouses are tomatoes and strawberries, respectively (Aguilar et al., 2020). Almería is within tile T30SWF.

Recent greenhouse mapping applications focused on the Anamur district in Turkey (Balcik et al., 2020; Zhang et al., 2022). There, the grown banana and strawberries supply 40% of the national demand (Blacik et al., 2020). In order to have a study site close but not coincident with the site used for calibrating the location-specific threshold for the APGI, we study the surroundings of the city of Demre from tile T35SQA.

For testing the robustness of OPAC over not yet studied areas but with similar horticultural models and landscapes of the above mentioned Mediterranean study sites, we visually identified in Google Earth two sites located in the surroundings of Marina di Acate (Sicily, Italy) within tile T33SVA and of the Ierapetra district (Crete, Greece) within tile T35SLU.

2.2. Data sources

2.2.1. Sentinel-2 satellite images

We use Sentinel-2 satellite images (L2A products) downloaded from the Copernicus Open Access Hub. Each product is a 100 × 100 km^2 ortho-image in UTM/WGS84 projection, extracted from a single orbit. The Sentinel-2 mission comprises a constellation of two identical satellites in the same orbit, 180° apart providing a global revisit time of maximum 5 days. The spatial resolution of their Multi-Spectral Instruments (MSI) ranges between 10 and 60 m. Sentinel-2A was launched on June 23, 2015 and Sentinel-2B on March 7, 2017; these dates indicate the limit for possible retrospective analyses. Sentinel-2 L2A products are processed with Sen2Cor (Main-Knorn et al., 2017), which uses Level 1C Top-Of-Atmosphere (TOA) input, performs its atmospheric, terrain and cirrus correction, and outputs bottom-of-atmosphere (BOA) reflectance multiplied by 10,000. In our analyses, we referred to the scaled BOA reflectance as reflectance. L2A products comprise a SCL with 20 m resolution pixel flags for: NO_DATA, SATURATED_OR_DEFECTIVE, DARK_AREA_PIXELS, CLOUD_SHADOWS, VEGETATION, NOT_VEGETATED, WATER, UNCLASSIFIED, CLOUD_MEDIUM_PROBABILITY, CLOUD_HIGH_PROBABILITY, THIN_CIRRUS and SNOW.

2.2.2. Very high-resolution images

Ground truth (annotated reference data) is essential for training machine-learning models. For Switzerland, we use freely available, very high-resolution (10 cm) aerial images with known acquisition dates for accurate interpretation of Sentinel-2 True Color Images (TCI). These aerial photos are updated every three years. Depending on the location, they are taken in different periods of the year to achieve the best image quality. These images can be downloaded from https://www.swisstopo.admin.ch/en/geodata/images/ortho/swissimage10.html; they are also available via the ESRI Satellite-Quick Map Service of QGIS. The shapefile of the flights with the corresponding date can be downloaded from http://data.geo.admin.ch/ch.swisstopo.lubis-bildstreifen/ch.swisstopo.lubis-bildstreifen.zip.

For study areas outside Switzerland, we infer the land cover captured at Sentinel-2 acquisition dates from freely accessible Google Earth imagery and Google Street View scenes.

2.2.3. Agricultural land cover

According to the Food and Agriculture Organization (FAO) of the United Nations, agricultural land refers to the share of land area that is arable (temporary crops, temporary meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow), under permanent crops (flowering shrubs, fruit trees, nut trees, and vines) and land used for forage. We acquire two vector datasets defining agricultural lands in Switzerland.

The first vector dataset is compiled by Swiss Cantonal Authorities and defines agricultural lands at the parcel level (*Landwirtschaftliche Bewirtschaftung* in German, hereafter LWB; available at https://www.geodienste.ch/downloads/lwb_nutzungsflaechen). It is updated at best every year depending on the canton. Overall, there exist 152 land covers annotated in the three Swiss national languages. Some of these land covers clearly indicate the presence of protected agriculture, e.g. in German "Gemüsekulturen in geschütztem Anbau ohne festes Fundament" or translated in English "Horticulture performed under cover without fixed foundations". Note that, in a few instances, we observed that more than one crop was grown in the same polygon of the yearly highly-detailed LWB vector dataset. According to the LWB vector dataset, the Swiss agricultural landscape was divided in 1,275,178 parcels for an overall area of 1,028,253 ha, similar to the area reported in the official report by UST (2021) (Section 2.1). A considerable percentage of parcels in the Swiss agricultural landscape are smaller than the resolution of the Sentinel-2 Band 1 equal to 60 m × 60 m. For example, Fig. 2 shows that about 80% of the parcels covered by all-year protected crop are smaller than 3600 m^2. Land uses within the categories of "Agriculture land" (e.g., strawberries), "Orchard" and "Other perennial culture" (e.g., perennial berries), but not limited to, can be covered by plastic mulches and greenhouses.

The second vector dataset is the Corine Land Cover map (hereafter Corine) for 2018, which defines agricultural lands at the EU scale (Table 2). Corine can be downloaded upon free registration from https://land.copernicus.eu/pan-european/corine-land-cover/clc2018. Corine has a minimum mapping unit of 25 ha, which is large relatively to Switzerland's small-scale farming, and cannot be used for emulating a segmentation of the land use and avoiding mixed spectral signatures.

3. Methods

The random forest-based Open field and Protected Agriculture land cover Classifier (OPAC) is fully developed in the RStudio environment (v4.1.1). Random forest is a machine-learning method, developed by Breiman (2001), which can be used for classification tasks. It uses "bootstrapping" and "feature bagging" to fit a user-defined number of individual decision tree models to the dataset; then, each tree acts as a classifier on its own and the final class assignation corresponds to the mode of the outputs. "Bootstrapping" is realized by randomly sampling the dataset with replacement for building each decision tree. "Feature bagging" is obtained through the random selection of a subset of the features present in the dataset at each split in the growth of each decision tree. Inherently, a random forest model contains several parameters. In this research, we fine-tune (1) the number of features randomly sampled as candidates at each split (mtry), (2) the number of decision trees to grow (ntree) and (3) the minimum terminal node size (nodesize). In the next subsections, we describe the routine followed to build the dataset used to develop OPAC.

Here, we focus on eight land covers (plus one class for shadows) that are predominant on Swiss agricultural lands and are also present in SCL (except protected agriculture). This choice allows for investigating the differences between OPAC and SCL, which can contribute to enhance SCL by adding protected agriculture. Thus, we set our research as a nine class, per-pixel classification problem by focusing on the land covers listed in Table 3. We neglect the class for "clouds" given its frequent misclassifications with protected agriculture and vice versa all year round (output not shown). For completeness, we list here special labels that we use to annotate pixels where OPAC encounters peculiar misclassification issues outside the training data set (results in Section 4.3): "Excavation", "Soil_next" and "NA". "Excavation" indicates the occurrence of construction works in the agricultural parcels, which are clear in the aerial photos and look like very bright surfaces in Sentinel-2 TCI. "Soil_next" highlights the relevance of the poor spatial resolution of Band 1 for mapping protected agriculture by showing those pixels misclassified as protected agriculture and within a distance of 60 m (sensor resolution) from true protected agriculture pixels. The class "NA" is used for pixels that could not be labelled confidently in the available aerial photos, nor in the 10 m resolution Sentinel-2 TCI.

3.1. Pre-processing

We develop OPAC based on Sentinel-2 L2A data pre-processed to 20 m pixel resolution. For each Sentinel-2 L2A pixel, we form a 13-dimensional feature vector by stacking ten spectral bands (i.e, Bands 1,2,3,4,5,6,7,8,11,12) and three spectral indices. These indices are the well-known NDVI (Rouse et al., 1974) and NDWI (Gao, 1996) together with the recent NI (Zhang et al., 2022) (mathematical formulations in Table 4). The NI is the most important feature to separate

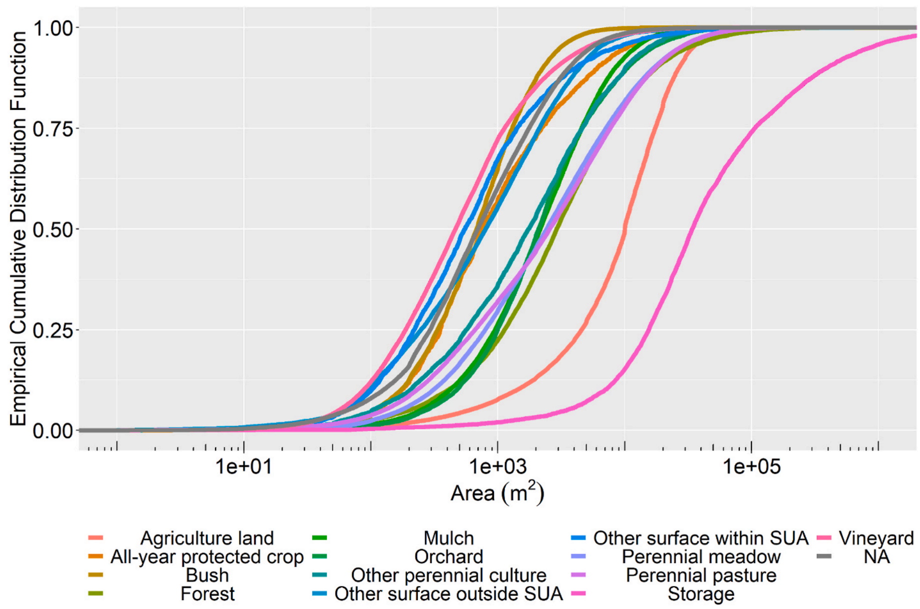

Fig. 2. Empirical cumulative distribution function of the areas of the parcels belonging to the agriculture land cover categories (different colors) defined by the Swiss Federal Office for Agriculture. SUA stands for Surface Useful for Agriculture. NA indicates the lack of the category definition for a land cover.

Table 2
Corine land cover code and label used in this study.

Corine Land Cover code	Label
211	Non-irrigated arable land
212	Permanently irrigated land
213	Rice fields
221	Vineyards
222	Fruit trees and berry plantations
223	Olive groves
231	Pastures
241	Annual crops associated with permanent crops
242	Complex cultivation patterns
243	Land principally occupied by agriculture with significant areas of natural vegetation
244	Agro-forestry areas

high-transparency plastic greenhouses from artificial surfaces, blue-roof building typical of the Chinese agricultural landscape, water and croplands (Zhang et al., 2022). We do not use APGI as a feature because APGI is the linear composition of NI with the red band (this band was used in the APGI to increase the separability with blue roof buildings) and with the aerosol band (inherently relevant for greenhouses mapping). Therefore, our working hypothesis is that the random forest model can learn the relationship (linear or not) among NI, the red band and the aerosol one from the data. Most of the instruments on Sentinel-2 operate at 20 m pixel resolution. Therefore, we down-sample Bands 2, 3, 4 and 8 from 10 m to 20 m (by means of the “aggregate” function and using the mean value) and we up-sample Band 1 from 60 m to 20 m (by means of the “disaggregate” function and keeping the same value). We disregard Band 8A but we keep Band 8 given that the latter is used to calculate the normalized difference water index (NDWI) (Gao, 1996), which we use as a model feature. We neglect Bands 9 and 10 because they are sensitive to water vapor absorption and cirrus clouds. We include three indices in the feature vector calculated as given in Table 4.

3.2. Model development

We describe the model development procedure in the next paragraphs. Fig. 3 shows the corresponding workflow.

Table 3
Land covers classified by OPAC.

Land cover label	Description	Notes
“Mulch_white”	White plastic mulch and while soil cover sheets.	We neglect black plastic mulch given that we are not able to distinguish them from other dark land covers in our preliminary investigations.
“Tunnel”	White low- and high-greenhouses in plastic.	Crops are mostly grown in the soil. Some level of automation can be realized within high-tunnels (e.g., sunlight shades). Hydroponics systems are possible.
“Glasshouse”	These are high-tech greenhouses, typically made from glass or polycarbonate, they look like large buildings with indoor vegetation or not depending on the plant life cycle and shades management.	These are fully automated greenhouses with hydroponics systems and sunlight shades.
“Vegetated”	Areas covered with low vegetation (crops, meadows, etc.) and tall vegetation (orchards).	Low and tall vegetation can have different spectral signatures. Among annual crops, there are summer and winter crops. The latter lead to the presence of vegetation in agricultural lands in winter and spring.
“Forest”	Areas covered with forests.	We found similarities in the spectral signature of orchards included in the class “Vegetated” with the spectral signature of trees in “Forest”; further investigations could improve the classification results.
“Baresoil”	Soil that is bare.	Limited to agricultural soils.
“Water”	Ponds, lakes, streams and rivers.	Sediments, algae and vegetation in water bodies can produce very different spectral signatures from water itself.
“Snow”	Snow.	Aging and dust can affect the spectral signature.
“Shadow”	Cloud and tree shadows.	

Table 4
Indices included in the feature vector.

Description	Formula
The Normalized Difference Vegetation Index relevant for mapping green vegetation (Rouse et al., 1974)	$NDVI = \frac{(B8 - B4)}{(B8 + B4)}$
The Normalized Water Moisture Index "NDWI" (Gao, 1996) relevant for assessing vegetation water content	$NDWI = \frac{(B8 - B11)}{(B8 + B11)}$
A recent and robust index for good separability of transparent plastic greenhouses from other natural and artificial land covers (Zhang et al., 2022)	$NI = \frac{(2 \times B8 - B4 - B12)}{(2 \times B8 + B4 + B12)}$

3.2.1. *Pixel labelling*

The final dataset used to develop OPAC comprises the information gained from labelling 5151 pixels of selected Sentinel-2 products (number of pixels per acquisition date and location of the Sentinel-2 image are shown in Fig. 4) using the eight land covers (plus a class for shadows) listed in Table 3. The labelling is based on the combined visual analysis of 10 m Sentinel-2 TCI and the available very high-resolution aerial images. Google Street View images are very helpful for acquiring the side view of the object framed in the pixel. Asynchrony among images is very likely and may hinder the confident classification of a pixel; when the latter happens, we do not use the pixel for labelling.

During model development, we labelled more than 6000 pixels, eventually filtered to 5151 after considerations of the reflectance values for Band 1 of the classes "Tunnel" and "Glasshouse". The pixels are sampled within two UTM grid squares in Switzerland (T32TLT and T32TNT) and one square in Spain (T30SWF) (tiles with red contour in Fig. 1) spanning from February to December over 2019 and 2021. This procedure is key to capture spatial and temporal heterogeneities in the dataset relevant for land cover mapping of agricultural areas. The UTM grid squares for Switzerland are predominantly covered by agriculture (open field and protected), they encompass diverse meteorological and hydrological regimes, topographic gradients, and they capture possible different crop management practices due to the influence of two cultural backgrounds, French and German (Fabre et al., 2021). We include 102 annotated pixels in the UTM grid square T30SWF of the Almeria region within the summer period (August 28, 2021) (Fig. 1) for achieving the transferability of OPAC to Mediterranean agricultural landscapes.

3.2.2. *Cross-validation of the dataset*

The iterative size increase of the dataset grounds on a quantitative and a qualitative assessment of the classification outputs calculated for several Sentinel-2 products upon using a range of values of the hyper-parameter space (Fig. 3). The parameter mtry, per default equal to the square root of the total number of predictors, was changed from 2 to 8, while ntree, per default equal to 500, was changed from 100 to 2000; nodesize was kept to 1. The quantitative assessment considers classification performance scores. The performance scores are calculated upon performing a 5-fold cross-validation for EQ (1) to EQ (5) with *TP* being the number of true positives, *TN* the number of true negatives, *FP* the number of false positives and *FN* the number of false negatives:

$$Overall\ accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad \text{EQ 1}$$

$$Precision = \frac{TP}{(TP + FP)} \quad \text{EQ 2}$$

$$Recall = \frac{TP}{(TP + FN)} \quad \text{EQ 3}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{(Precision + Recall)} \quad \text{EQ 4}$$

$$Intersection - Over - Union = \frac{TP}{(TP + FP + FN)} \quad \text{EQ 5}$$

In the qualitative assessment, we run OPAC on Sentinel-2 products within the UTM grid squares selected in this study (tiles with the contour in black or red in Fig. 1) at random acquisition dates. Then, we visually compare the classification map against the array of reference images available (i.e., 10 m Sentinel-2 TCI and very-high resolution ones). Note that, we run the model only on Sentinel-2 pixels completely within the vector polygon features (no intersection with polygons' boundaries) of LWB and Corine (within Switzerland) and of Corine (outside Switzerland). In this way, we focus on agricultural lands and we reduce the possibility to classify pixels with a mixed land cover. However, there is the possibility that more than one crop is grown in the same polygon of LWB, which could result in a mixed spectral signature.

3.2.3. *Dataset quality check relative to band 1*

In the final dataset, we filtered out pixels labelled as "Tunnel" and "Glasshouse" with Band 1 reflectance <1000. Zhang et al. (2022) demonstrate that high reflectance values for Band 1 are more likely than other bands to indicate plastic greenhouses. We confirmed this observation by plotting the reflectance values for Band 1 of the different land covers over the month of the year (MOY). This analysis allowed to identify a reflectance threshold of 1000 for Band 1 for separating "Tunnel", "Mulch_white" and only in part "Glasshouse" from other land covers (Fig. 5). However, Band 1 has a poor pixel resolution (60 m), which may have led to low (<1000) Band 1 reflectance values of some labelled pixels due to mixed signature. Yet, the reflectance values of Band 1 for "Glasshouse" are significantly higher and they feature seasonal variability in the West of Switzerland, but not in the East. The range of Sentinel-2 observation geometries is however quite comparable, and we cannot explain this observation without dedicated inspections.

We investigate the implications of training a model on the full dataset or on the final dataset (having filtered out pixels labelled as "Tunnel" or "Glasshouse" with a reflectance value < 1000 for Band 1). First of all, we highlight how blurred the information can become when the pixel resolution decreases from 10 cm down to 10 m, 20 m and 60 m, which is the one of Band 1 (Fig. 6a–d). Second, we show that the filtered dataset allows for developing a more accurate version of the model (i.e., OPAC), for example, given the capability to classify as "Baresoil" rather than "Glasshouse" the mixed signature of bright surfaces with vegetation typical of roads with adjacent vegetation (Fig. 6e and f). By training the model on the filtered dataset (i.e., OPAC), we reduce the occurrence of false positives for the task of mapping protected agriculture. In the filtered dataset, we disregard 325 pixels labelled as "Glasshouse" and 113 pixels labelled as "Tunnel" that have reflectance values lower than 1000 for Band 01. However, excavation sites characterized by high reflectivity are still misclassified as protected agriculture (Fig. 6e and f). By constraining the model with a highly detailed vector dataset of agricultural lands (i.e. LWB), we can better exclude other land covers that can be confused with protected agriculture, such as highly reflective surfaces and roads (Fig. 6e–f against Fig. 6g–h).

3.3. *Cross-validation of the training dataset*

We quantify how well the model generalizes to a Swiss region and on several MOY of one year not used for model training. For this, a model is trained on all the labelled pixels in the UTM grid squares T32TLT and T32TNT for 2019 and 2021. Then, it is validated on the UTM grid square T32TMT using one product per month from February to November in 2020. We select products with a cloud cover percentage lower than 20% according to SCL. This cloud filter reduces the number of available T32TMT products from 113 to 31. Then, we keep the one product per MOY that visually contain the least clouds over agricultural areas, and we manually mask out remaining clouds from further analyses. For each MOY, we randomly generate a maximum of 15 pixels per assigned land cover from unique LWB polygons (to avoid repetitions), and manually label them based on the array of images available (i.e., aerial, from

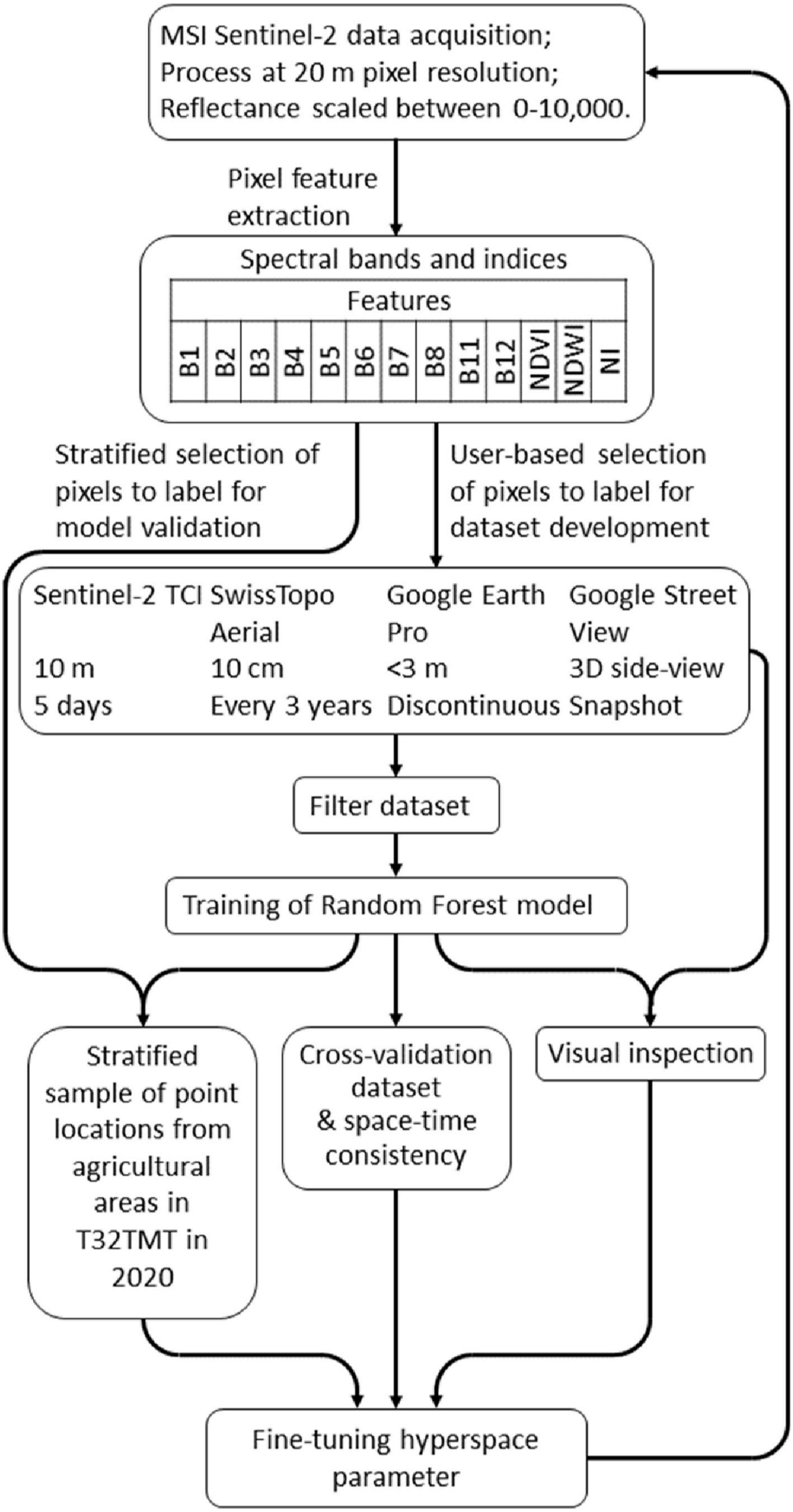

Fig. 3. Dataset and OPAC development workflow. We acquire Sentinel-2 L2A products. Bands reflectance values are scaled between 0 and 10,000. We resample bands spatial resolution to 20 m. We use Sentinel-2 True Color Images at 10 m for visual inspection. We extract reflectance values and we calculate indices; these constitute the model features. We export into our dataset the features from selected Sentinel-2 pixels and label them upon using the available imagery sources. We disregard from the dataset pixels with labels equal to “Tunnel” or “Glasshouse” and reflectance values for Band 1 smaller than 1000. We train OPAC on the dataset and we use it to classify Sentinel-2 products. We perform a cross-validation of OPAC training on the dataset. We carry out scenarios of spatial and temporal consistency (results in the Supporting Information). With focus on the monthly least clouded product for tile T32TMT over the year 2020, we generate a stratified sample with up to 15 point locations per classified land cover, which are used for quantitatively validating the generalization of OPAC. We visually compare the classified Sentinel-2 products against the available imagery sources for a qualitative assessment. Finally, the optimal model parameters are chosen. We opportunely select and label as many pixels as necessary until we satisfy the quantitative and qualitative tests.

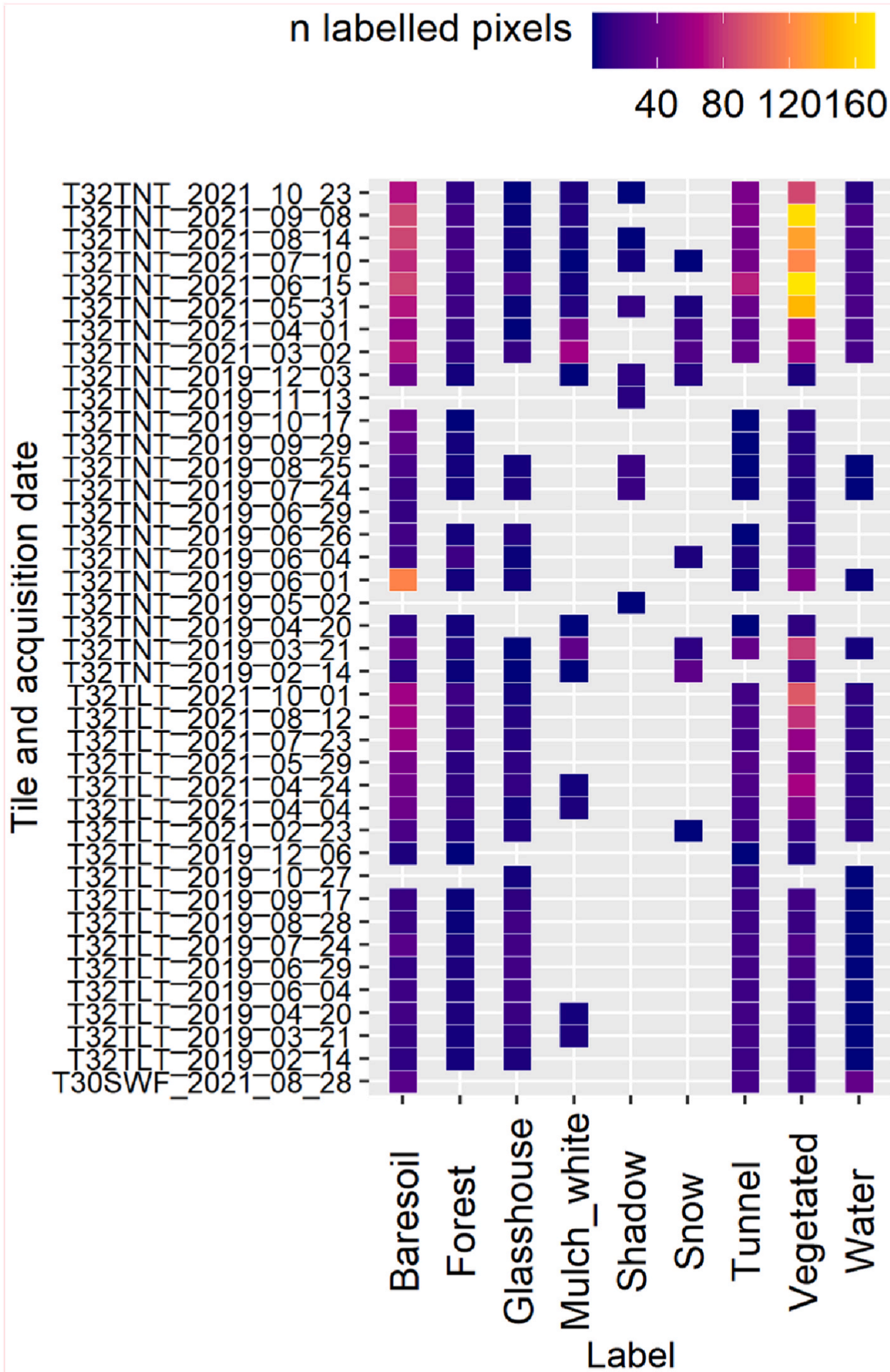

Fig. 4. Heatmap showing the number of labelled pixels by class (x-axis) distinguished by Sentinel-2 UTM grid square and acquisition date (y-axis) comprising the final dataset. Transparent tiles indicate no label was assigned.

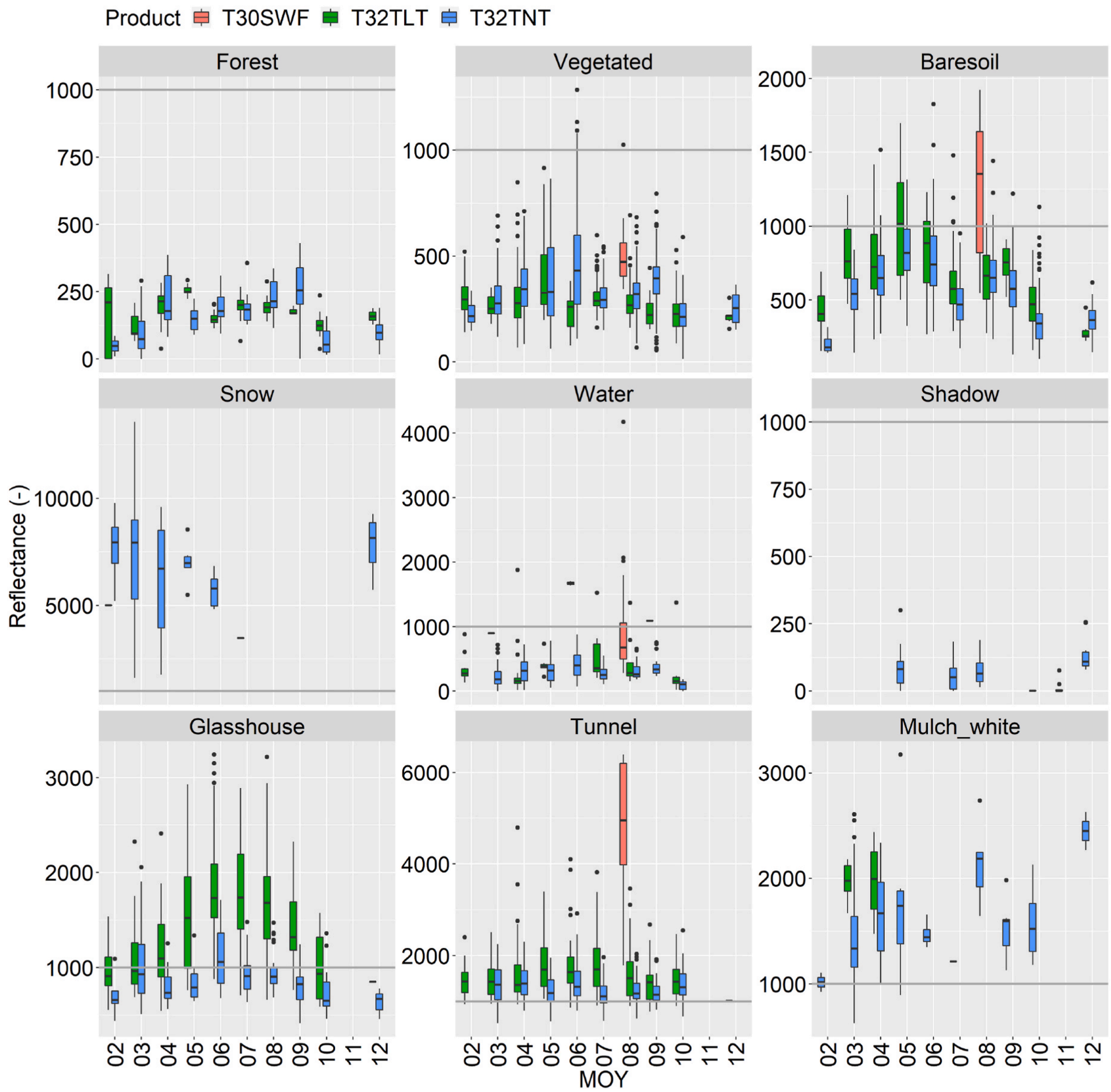

Fig. 5. Reflectance (BOA reflectance × 10,000) of Band 1 (y-axis) grouped by month of the year (MOY) and land cover (x-axis). The different products (T30SWF, T32TLT and T32TNT) are depicted with different colors. The horizontal grey line highlights the reflectance threshold at 1000.

Google and from Sentinel-2). For the random selection, we filter out Sentinel-2 pixels that have a sum of the coefficient of variation (CV) for the 10 m bands (Bands 2–4 and 8) larger than 20. This helps to avoid the selection of mixed land covers and it is particularly useful if one works with less detailed agricultural vector data (i.e., Corine).

Additional scenarios run to prove the consistency of the dataset over space and time are detailed in the Supporting Information.

3.4. *Assessment of the land cover assignments among classification methods*

We assess how pixels manually labelled in the scenario of Section 3.3 are classified by different methods: OPAC, APGI and SCL. OPAC and APGI are both pixel-based methods. However, OPAC is built on the Random Forest Machine Learning model and APGI is currently the best performing index that requires a space- and time-specific threshold for classification. A comparison with other published pixel-based classification models for plasticulture mapping is hindered by the unavailability of the corresponding training sets and trained models in public repositories. SCL does not have a class for protected agriculture and it is important to understand what classes are affected by protected agriculture (e.g., unaware users may consistently neglect these pixels in their scopes when programmatically relying on the SCL cloud mask to filter out clouds in Sentinel-2 L2A products).

3.5. *Model generalization to other countries*

As a transferability study, we run OPAC on Corine-delimited agricultural lands on four products in the Mediterranean area (i.e., Greece, Italy, Spain, Turkey), where plastic tunnels are widely used, and on one product in a temperate climate (i.e., the Netherlands), where high-tech greenhouses are common. We select products in which the Corine-delimited area is visually cloud-free (but not according to SCL). These include T30SWF in Spain on August 28th' 2021, T33SVA in Italy on July 16th' 2021, T35SLU in Crete on July 31st' 2021, T35SQA in Turkey on August 27th' 2021 and T31UET in The Netherlands on September 7th' 2021. For classification using APGI, a Sentinel-2 product-specific threshold must be defined. For this purpose, in each case study, we increase the APGI threshold from 0.2 with steps of 0.01 until the area of plastic greenhouses found by APGI matches the area of protected

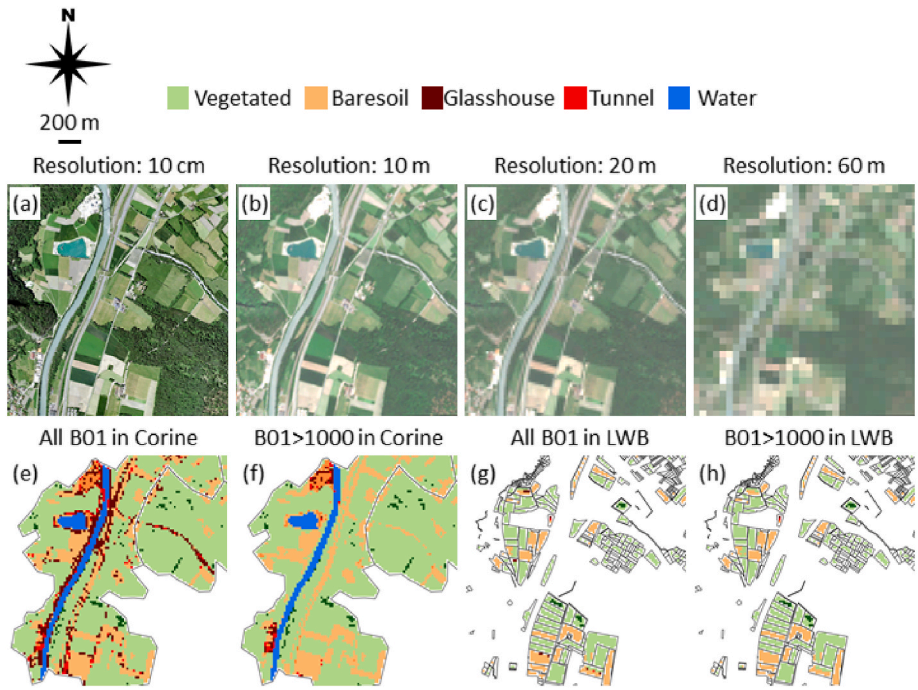

Fig. 6. Comparison of two versions of OPAC constrained by Corine and LWB (boundaries of the polygons as thin black lines), for a subset of T32TNT on June 1st, 2019. Very high-resolution aerial photo (a), CI from Sentinel-2 at 10 m (b), 20 m (c) and 60 m (d). (e) Training with full dataset, run within Corine; (f) training with filtered dataset, run within Corine. (g) Training with full dataset, run within LWB; (f) training with filtered dataset, run within LWB. Panels (g) and (h) have fewer classified pixels because we disregard agricultural parcels with an area smaller than the model resolution to avoid mixed land covers within a pixel (20 m × 20 m).

agriculture classified by OPAC. We depict the classification match, and mismatch, between OPAC and APGI. That is, we distinguish pixels where (i) both classifiers do not assign protected agriculture and where (ii) both, (iii) OPAC but not APGI, (iv) APGI but not OPAC assign protected agriculture.

4. Results

4.1. Signature land covers and their evolution over time

We explore the spectral characteristics of the manually labelled Sentinel-2 pixels (Section 3.2) to figure out the separability of the land covers in Table 3. The spectral signatures of vegetated land cover classes feature a maximum at near-infrared wavelengths that are representative of biomass (red edge) (Fig. 7). The seasonal red edge of “Forest” is much more prominent than for “Vegetated”; very likely, the latter land cover encompasses summer and winter crops, which together cancel out the effect of seasonality in biomass production. “Snow” reflectance signatures vary with age and dust content. For these reasons, “Snow” has smaller reflectance values for Bands 1–8 and higher values for Bands 11 and 12 in summer than in winter. The reflectance signature of “Water” is generally characterized by low values and it changes with the type of water body (i.e., alpine lake versus pond or alpine stream versus lowland river), primary productivity and sediment transport (Odermatt and Gege, 2022). “Shadow” is also characterized by low reflectance values; we note higher values in the red edge spectrum representative of biomass, which may be sensed although shadowed by clouds or other obstacles like tall trees. “Baresoil” has a predominance in the short wave infrared spectrum due to minerals, such as clay and carbonates (Rowan and Mars, 2003). The land covers included in protected agriculture are high-tech greenhouses (i.e., “Glasshouse”), plastic tunnels (i.e., “Tunnel”) and white plastic mulch (i.e, “Mulch_white”). The average signature of protected agriculture follows the typical pattern of the class “Vegetated” but with higher reflectance for Bands 1–4 and lower reflectance for Bands 5–8. This could be due to the brightness conferred by white plastic and the presence of wavelength selective pigments in the chemical composition of the cover (Mormile et al., 2017). Here, we report for the first time the signature of high-tech greenhouses (“Glasshouse”) captured by Sentinel-2. The signature of “Glasshouse” is similar to the ones of “Tunnel” and “Mulch_white” but generally shifted to lower reflectance values across all bands. Both “Tunnel” and “Glasshouse” can be characterized by substantially high reflectance across all bands. This very likely corresponds to the use of shade covers to protect the plants from intense sunlight as well as to reduce heat gain inside the greenhouse. Among all classes, “Tunnel” shows the largest monthly variability in reflectance. This may be attributed not only to the phenology of the crop (Nemmaoui et al., 2018) but also to the material used for different tunnels sampled in the same MOY (e.g., polyethylene, polyvinyl carbonate, polycarbonate, etc.) as well as other properties and practices that can affect optical properties of plastic greenhouses (Mormile et al., 2017).

4.2. Cross-validation of the training dataset

The final training dataset is composed of 5151 labelled pixels from the products T30SWF, T32TLT and T32TNT in the years 2019 and 2021 (Fig. 4). In the running version of OPAC, mtry is equal to 2, ntry is equal to 200 and nodesize to 1. We find that a better training set rather than an extensive hyper-parameter space tuning results in a more robust model. The 5-fold cross-validation of the final dataset reveals a constant overall accuracy of about 92–93% over the 5 runs. Fig. 8a depicts the confusion matrix obtained from the 5 runs where we show the sum of the number of matches between the labelled class (x-axis) and the predicted class (y-axis) over the 5 runs. It highlights the number of matches among pairs of classes normalized by the number of pixels within each labelled class. Evidently, different classes of protected agriculture (i.e., “Mulch_white”, “Tunnel” and “Glasshouse”) are likely to be misclassified among each other. The poor performance in distinguishing among these classes is not unexpected, given that their in-class spectral similarity is low in relation to their across-class dissimilarity.

We plot the mean decrease in Gini coefficient that is one measure to understand the feature importance for best splitting of the data at the node. A high mean decrease in Gini coefficient indicates that the feature is important. This is because the omission of the feature will result in trees with nodes that contain elements of the dataset from multiple labelled classes rather than a single one. In our research, the features NDVI, NDWI, Band 1, Band 2 and NI are the most important five features (Fig. 8b).

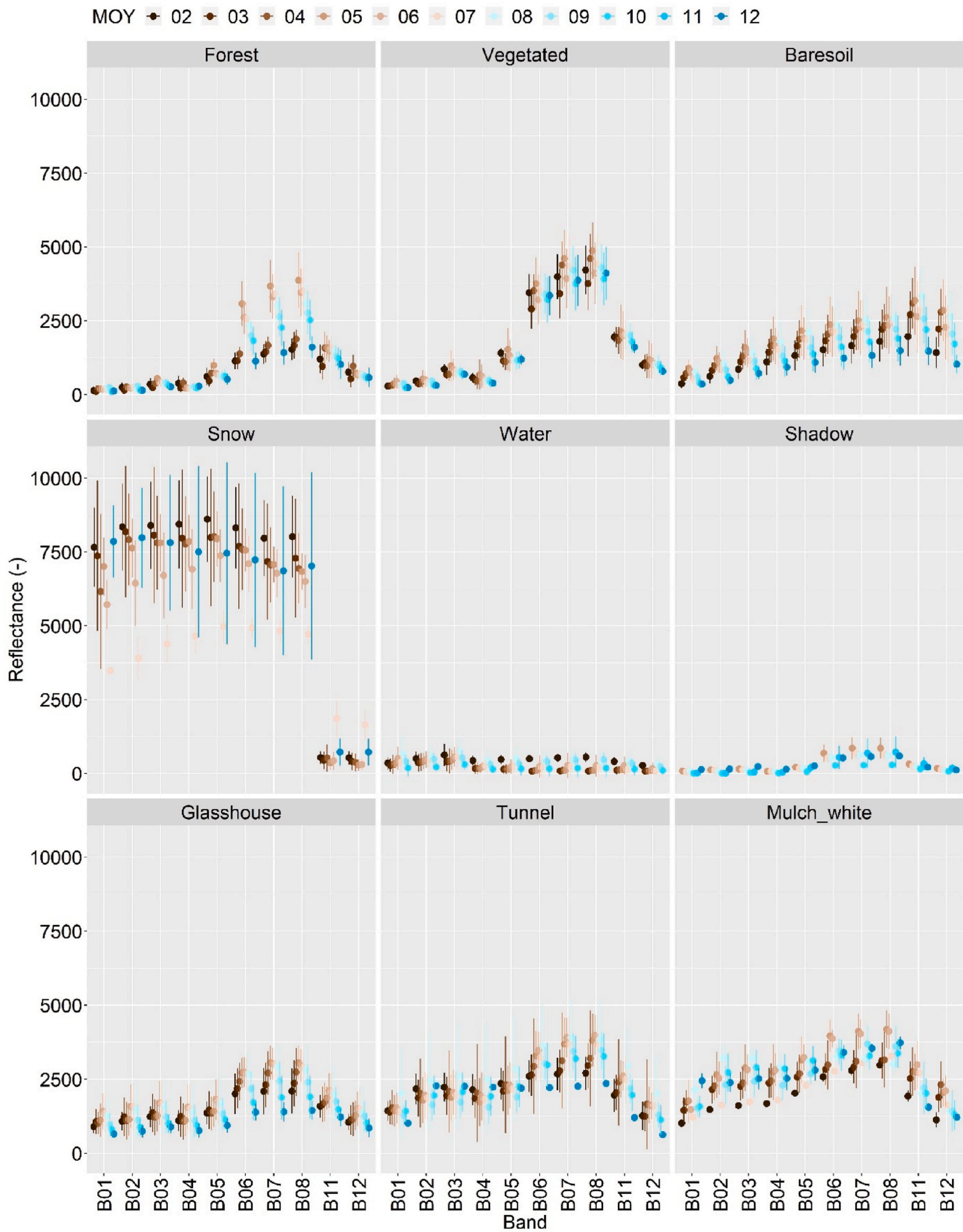

Fig. 7. In the y-axis, we show the mean reflectance (BOA reflectance × 10,000) (filled circles) and the standard deviations (colored vertical bars) for each month of the year (MOY; filling colors) shown sequentially for each band along the x-axis per land cover. Means and standard deviations consider all pixels used to train OPAC (UTM grid squares T32TNT, T32TLT and T30SWF, and years 2019 and 2021).

The performance scores increase up to 98–99% when we group the three classes of protected agriculture together. We calculate other performance scores for a fair assessment of the model, especially when class imbalances can play a role (Table 5). These scores in the confusion matrix corroborate satisfying classification capabilities for all land covers but the ones of protected agriculture, and it highlights misclassification issues that can largely be related to spectral ambiguity. For example, we know that both "Tunnel" and "Glasshouse" can have similar spectral signatures, such as (i) high reflectance across bands when measures are likely taken to reduce the incoming sunlight and (ii) vegetation-like signature but brighter given the higher reflectance in Bands 1–4. Accordingly, about 40% of all pixels labelled as glasshouses were accidently classified as tunnels.

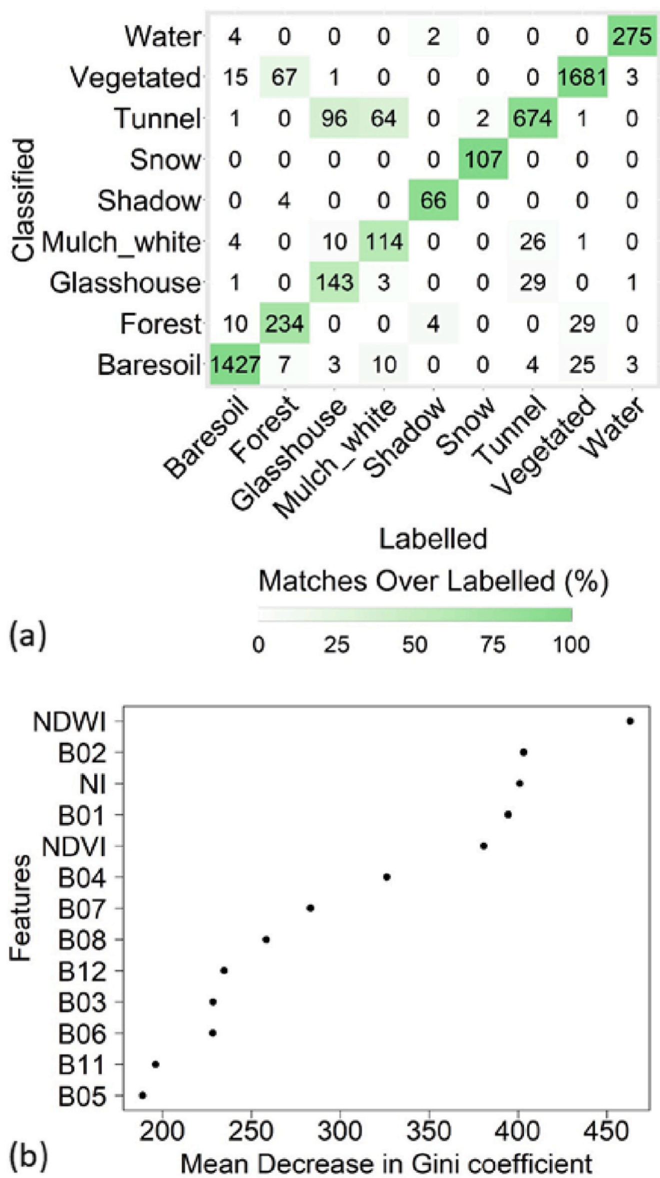

Fig. 8. Classification accuracy and features importance analysis. (a) Confusion matrix of the 5-fold validation on the final model. Numbers in the cells represent the number of matches between the labelled class (x-axis) and the predicted class (y-axis) over the 5 validation runs. The background color depicts the number of matches normalized by column (i.e., the number of pixels per each labelled class) over the 5 runs. (b) Mean decrease in Gini coefficient for the features in OPAC.

4.3. Scenarios of spatial and temporal consistency

We assess to what extent the final dataset for Switzerland (obtained from our final dataset, excluding labelled pixels from T30SWF on August 28, 2021 as justified in Section 3.3) can robustly classify an agricultural area in a different region of Switzerland (T32TMT) and in the MOY of a different year (2020) under the hypothesis that the model is sufficiently trained over space (T32TLT and T32TNT) and time (2019 and 2021). OPAC distinguishes between protected and open field agriculture and it still suffers to separate the three classes of protected agriculture (i.e., “Glasshouse”, “Mulch_white” and “Tunnel”). In fact, the differences between the overall accuracies calculated after grouping the three classes of protected agriculture as one and the overall accuracies calculated using each land cover separately can range from 11% (in May and August) up to 30% (in April). The overall accuracies for the months of May and June are decreased by high misclassification rate between “Water” and “Shadow”, which both appear as dark objects (low reflectance across bands). By counting the number of pixels labelled as “Mulch_white” in Fig. 9, it becomes evident that this land cover is predominant in March and April; this is expected because plastic mulch serves to increase soil temperature and control weeds during the early stage of crops growth, among other uses. After that, the growing crops can hide the plastic mulch beneath from optical sensing. From June on, we labelled a larger abundance of “Tunnel”; again, this is not a surprise because plastic covers protect ripening fruits from adverse weather conditions, such as summer storms and hail.

In this analysis, we encounter few interesting misclassification issues, which we aim to address in future works. Rather than sampling another pixel for manual labelling using the list of land covers in Table 3, we assign new labels as specified in Section 3.2, which are: “Excavation”, “NA” and “Soil_next”. This approach reveals cases where the model can fail. “Excavation” indicates the occurrence of construction works in the agricultural parcels, which are clear in the aerial photos and looking as very bright surfaces in Sentinel-2 TCI. Given the high reflectivity, very likely due to high dryness of the bright soil, these pixels are often misclassified as “Glasshouse” or “Mulch_white” rather than “Baresoil”. The class “NA” indicates that from the available aerial photos and the 10 m resolution Sentinel-2 TCI it is not possible to confidently label the pixel. Rather than sampling another point in this case, we keep this information to stress the importance of very high-resolution images or ground truth observations for accurately labelling a land cover. “Soil_next” is a crucial class for validation purposes as it indicates the relevance of the spatial resolution of Band 1 for mapping protected agriculture. This label indicates pixels of bare soil or of low and sparse vegetation within 3 pixels distance from surfaces highly reflective in Band 1 (the resolution of Band 1 is 60 m and we are using a 20 m pixel resolution). Therefore, surfaces that are highly reflective in Band 1 can lead to the misclassification of nearby pixels. In addition, low- and high-tech greenhouses can be realized with water retention ponds to store rainfall water intercepted by the greenhouse roof. The presence of water storages next to plastic greenhouses would contribute to the mixed land cover within a pixel issue.

Table 5
Tabulated classification performance scores. The scores for the land cover in bold “**Protected agriculture**” in the last row are recalculated after grouping together the classes “Glasshouse”, “Mulch_white” and “Tunnel” in both the labelled and predicted classes.

Land cover	*Precision*	*Recall*	*F1-Score*	*Intersection-Over-Union*
“Baresoil”	0.96	0.98	0.97	0.94
“Forest”	0.84	0.75	0.79	0.66
“Glasshouse”	0.81	0.57	0.67	0.5
“Mulch_white”	0.74	0.6	0.66	0.49
“Shadow”	0.94	0.92	0.93	0.87
“Snow”	1	0.98	0.99	0.98
“Tunnel”	0.8	0.92	0.86	0.75
“Vegetated”	0.95	0.97	0.96	0.92
“Water”	0.98	0.98	0.98	0.95
“**Protected agriculture**”	0.99	0.98	0.98	0.98

4.4. Comparison among classification methods

We describe in detail the validation of the classification output by OPAC against our manually labelled pixels for T32TMT in 2020 in the previous section. Here, we use alluvial plots to investigate how the different classification methods label the same pixels (Fig. 10). We use this approach to understand the information flow keeping in mind that OPAC, APGI and SCL have different classes. In particular, SCL does not

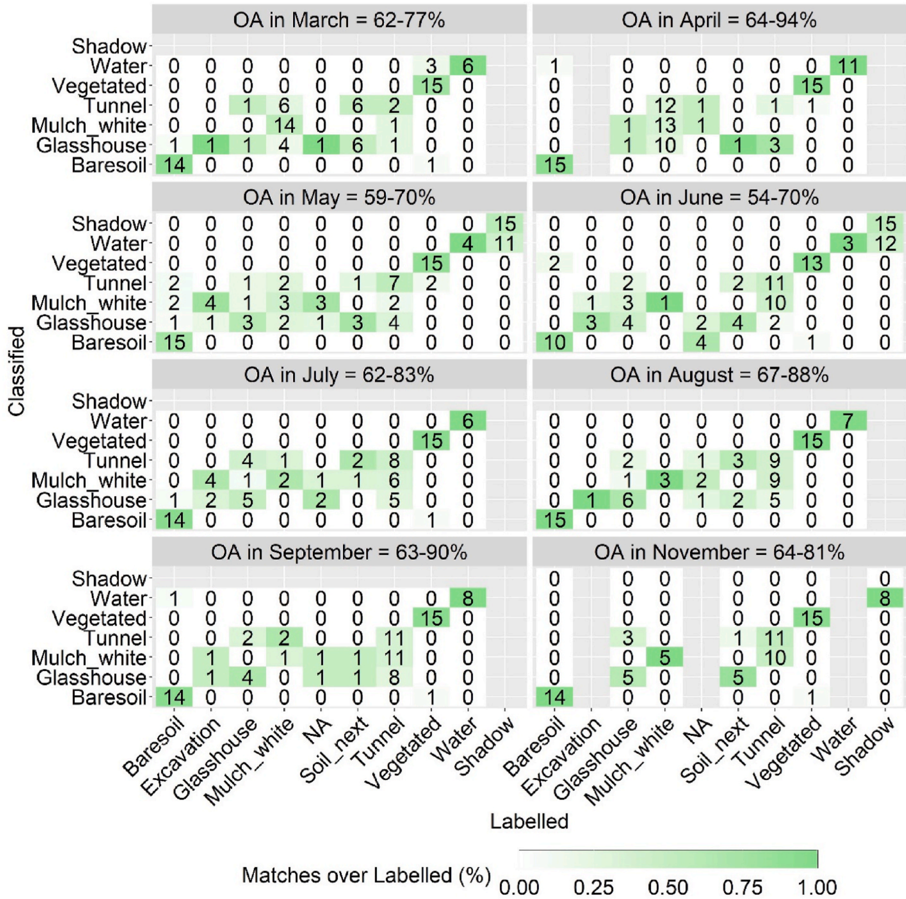

Fig. 9. Confusion matrices for each month of the year resulting from the validation of OPAC generalized to another product (T32TMT) in a different year (2020). Numbers in the cells represent the number of matches between the labelled class (x-axis) and the predicted class (y-axis). The background colors range from white to green and depicts the number of matches normalized by the number of pixels per each labelled class. The titles of the subpanels indicate the month of the year and the calculated overall accuracies (OA) considering all land covers and, separated by a dash, after grouping "Glasshouse", "Mulch_white" and "Tunnel" together as one unique land cover of "protected agriculture". Note that, the class "Forest" is renamed as "Vegetated" before sampling the pixels to be labelled. We labelled maximum 15 randomly sampled pixels, without repetition per parcel, per each land cover predicted by OPAC.

have a class for protected agriculture. We find that SCL flags protected agriculture as UNCLASSIFIED, NOT_VEGETATED, CLOUD_HIGH_PROBABILITY, CLOUD_MEDIUM_PROBABILITY, and, to a small extent, VEGETATION. OPAC and SCL have similar classes for vegetated and bare areas. "Excavation" is mostly related to NOT_VEGETATED areas but also to UNCLASSIFIED and cloud-related classes; this corroborates the difficult classification of such highly reflective bare surfaces. "Soil_next" relates to "Vegetated" and NOT_VEGETATED classes and in one instance it is related to cloud-related classes. The APGI is sufficiently capable of distinguishing between non-plastic surfaces and plastic agriculture. We find that it is able to capture a large percentage of "Tunnel" but less capable to map "Glasshouse" and "Mulch_white".

When we focus on the product T32TMT on August 9, 2020 (Fig. 10b) rather than year-long data, SCL flags protected agriculture as clouds and not vegetated areas or cannot classify the pixel. The APGI shows an improvement in the mapping of high-tech greenhouses and plastic mulch but it still misclassifies about 30% of pixels covered with plastic tunnels as in Fig. 10a. This improvement can be explained by the fact that the APGI needs a time-specific threshold for mapping plastic greenhouse, and the value of 0.41 used here, was determined for the Mediterranean climate for summer by Zhang et al. (2022).

4.5. Qualitative results: generalization to other countries

The overarching goal of land cover classification models is to be robust, so that they can be generalized at the country-, continental- and eventually the global-scale. Here, we visualize how OPAC and APGI perform in the Corine-delimited agricultural areas of other North- and South- European UTM grid squares (as explained in Section 3.5).

In Spain (T30SWF), on August 28th, 2021 (20210828), it is impossible to visually recognize the plastic greenhouses from Sentinel-2 TCI even at 10 m resolution due to the very high brightness of the surfaces, while large water reservoirs are distinguishable, and sometimes, the smaller water storages can be seen among the bright surfaces (Fig. 11a). Other non-agricultural soils or bare soils can be more or less easily recognized in the image. Most of the area is classified as protected agriculture by OPAC and 93% of these pixels are mapped as plastic greenhouses by APGI using a threshold of 0.28 (light blue pixels in Fig. 11b). This calculated threshold for APGI satisfies the condition that APGI maps an equivalent extent of plastic greenhouses as OPAC did for protected agriculture. OPAC maps additional pixels as protected agriculture in the area (red pixels) as well as APGI does for plastic greenhouses (yellow pixels); also, OPAC and APGI classifies pixels as not covered by protected agriculture or plastic greenhouses, respectively (grey pixels). To what extent OPAC and APGI are correct or not are assessed using the clear very high-resolution imagery from Google, with some uncertainty due to the asynchrony in the acquisition dates (Fig. 11c). OPAC misclassifies the small water storages as protected agriculture (sparse red single pixels among the light blue ones in the South of the shown area) as well as roads (the whole extent of the generated raster data are available in the online repository). In contrast, APGI classifies what look like bare soils as plastic greenhouses. Yet, a threshold of 0.41 rather than 0.28 was suggested by Zhang et al. (2022) for this Spanish region, which would result in: (1) a lower match between the two classifiers (less light blue pixels and more red pixels), (2) a decrease in the number of yellow pixels unless other currently mapped pixels as non plastic greenhouses will be classified as greenhouses.

In the Netherlands (T31UET), on September 9th, 2021 (20210907), the Sentinel-2 image shows several districts of high-tech greenhouses intertwined by canals, residential areas (Fig. 11d). In the East of the

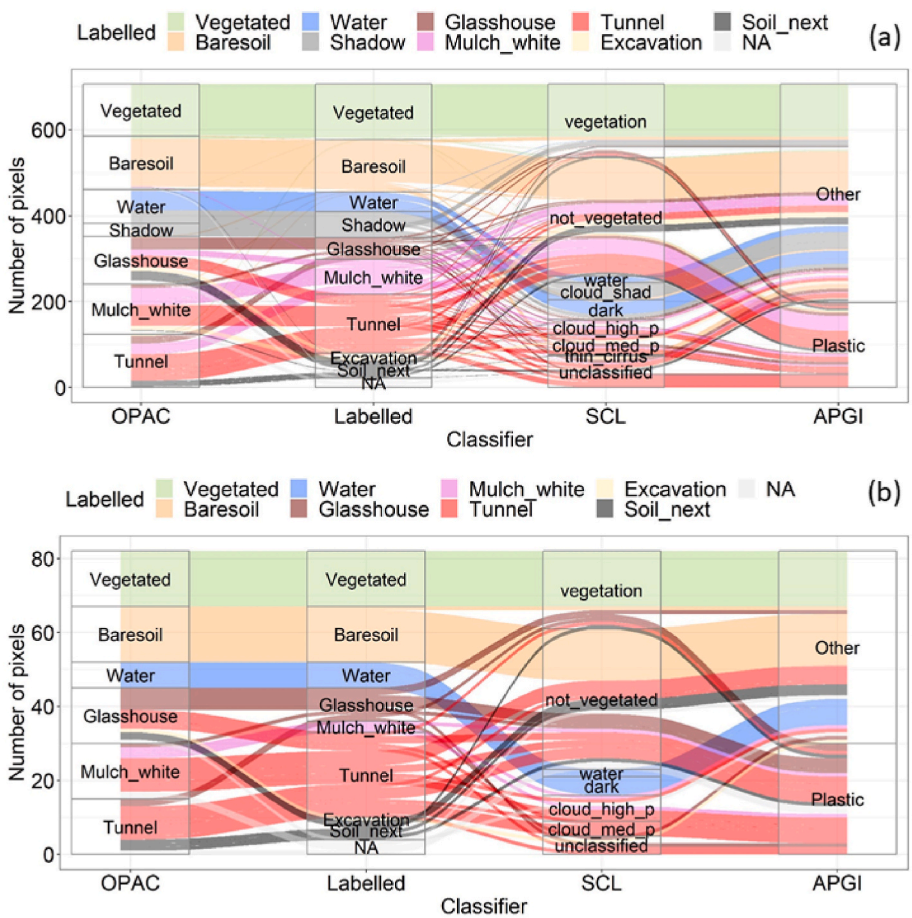

Fig. 10. *Alluvial plot showing the classification flow of the predicted land cover by OPAC, which is then manually labelled and compared against SCL and APGI for all months of the year 2020 in panel (a) and for the August 9, 2020 in panel (b). This order of the columns resulted in the clearest visualization of the classification flow. The height of the column per class is proportional to the number of pixels, labelled or classified, belonging to the class per input data. The width of the line corresponds to the number of pixels which class is matched between the two connected input data. SCL classes are changed to lower case for readability; "cloud_shad" is the abbreviation for "CLOUD_SHADOWS", "dark" for "DARK_AREA_PIXELS", "cloud_high_p" for "CLOUD_HIGH_PROBABILITY" and "cloud_med_p" for "CLOUD_MEDIUM_PROBABILITY". The colors of the strata refer to the labelled input data. For the APGI, we consider a threshold value for the index of 0.41 as recommended for Mediterranean areas in summer (Zhang et al., 2022).*

studied area, one could see intense light coming from a greenhouse, which may be due to some agro-management practices, such as soil disinfection by means of pyrolysis. Here OPAC accurately identifies more greenhouses as compared to APGI (Fig. 11e) (the calculated threshold and the match between the two classifiers is reported in Table 6). However, also APGI maps few additional greenhouses that OPAC does not, like in the South of the area. In these few cases, our previous version of a random forest model trained on the unfiltered dataset (including pixels with reflectance in Band 1 smaller than 1000) was able to classify these pixels as protected agriculture. Overall, it results that APGI is capable to map to some extent not only plastic tunnels but also high-tech greenhouses, which has not been demonstrated before.

In Italy (T33SVA on July 16th' 2021), Crete (T35SLU on July 31st' 2021) and Turkey (T35SQA on August 27th' 2021), the landscape observed through Sentinel-2 look similar to the one described for Spain (Fig. 11, panels g, j, m compared to panel a). However, the APGI threshold almost doubles for these Sentinel-2 products as compared to the one for Spain on August 28th' 2021, going from 0.28 up to 0.56 (Table 6). Moreover, the area mapped as plastic greenhouses by APGI ranges from 63% to 74% of the area classified as protected agriculture by OPAC (Table 6). Visually, OPAC can clearly detect more plastic greenhouses than APGI upon using the calculated thresholds. Interestingly, the highest number of yellow pixels (plastic greenhouses according to APGI only) are found in Crete (Fig. 11k), which appear to correspond to bare soils according to the Google imagery. Local knowledge would be needed to understand this classification output in order to improve the robustness of APGI.

5. Discussion

5.1. Relevance of reflectance value at wavelengths below 490 nm for protected agriculture mapping

OPAC is capable to classify multiple land covers from single Sentinel-2 products. Lands covered by open field or protected agriculture can be distinguished to a great extent thank to the aerosol band (Band 1) upon using a threshold of 1000. Indeed, Zhang et al. (2022) exploited Band 1 to develop the most recent index for mapping plastic greenhouses. However, the lower resolution of this band compared to the other ones (60 m against 10–20 m) hinders the full exploitability of this band in the highly spatially fragmented agricultural landscapes. In fact, a single plastic greenhouse can extend up to 100 m in length but typically only 10 m in width, although they could be attached side-by-side in some cases. Though, it is not clear, from a material science point of view, why this band is crucial for this mapping application. Zhou et al. (2022) carry out an in-depth investigation of the hyperspectral fingerprint of plastic greenhouses in the wavelength range from 970 nm to 2500 nm. That approach follows the statement by Schwanninger et al. (2011) that being plastics a byproduct of petroleum, its diagnostic absorption features are mainly observed in the SWIR region across the optical spectrum. In our research, we found that expanding the analysis of hyperspectral signatures down to 400 nm could be a promising area of investigation for improving the detection of plastics from satellites. However, we noted that some pixels belonging to "Tunnel" and "Glasshouse" have a reflectance in Band 1 lower than 1000. This is investigated to a small extent given the lack of ground-truth data. From Google Street View images, we noted that some bright structures from Sentinel-2 TCI and

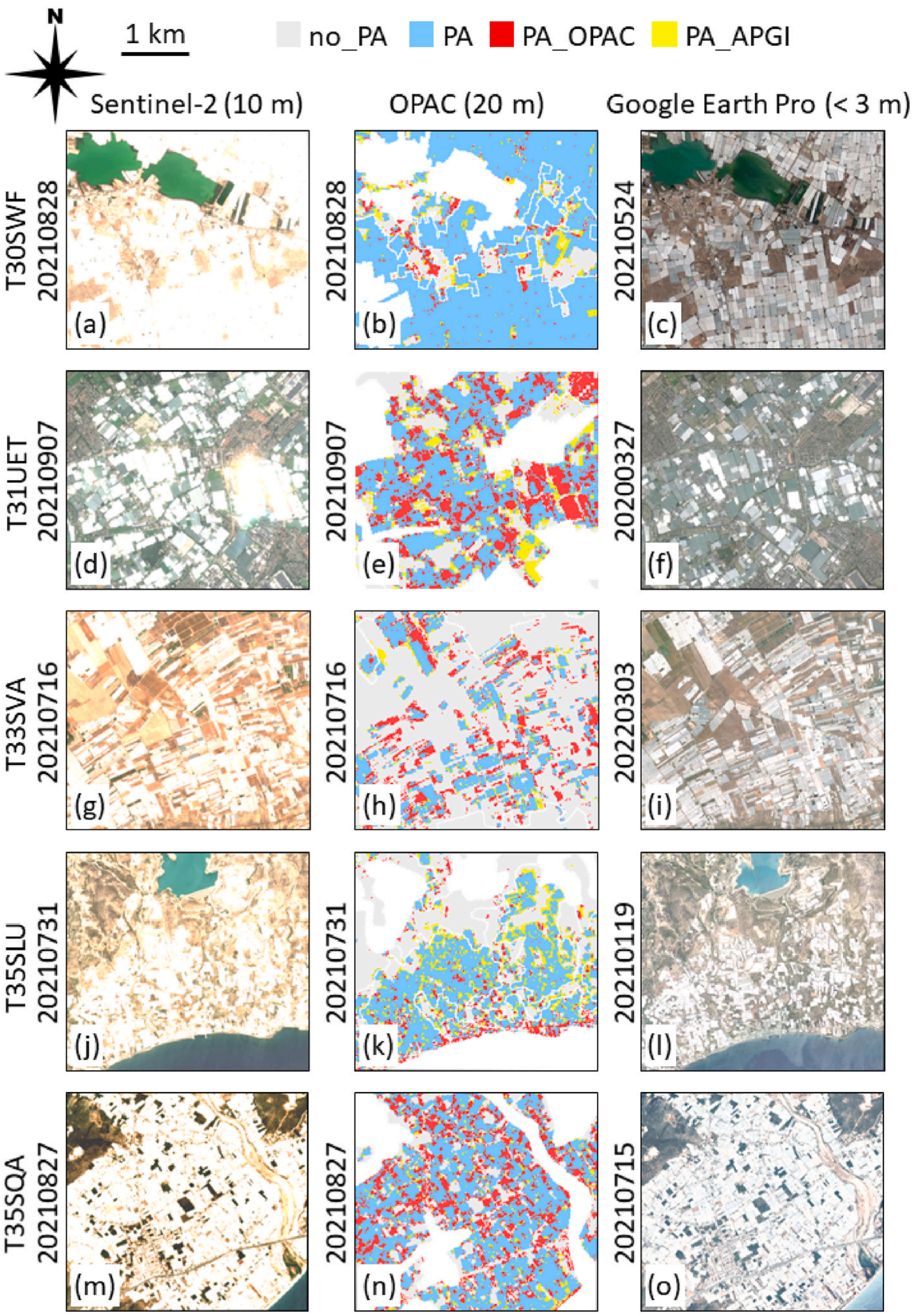

Fig. 11. Zoomed in visualization of the generalization assessment. Left panels: Sentinel-2 TCI at 10 m pixel resolution. Centre panels: comparison of the mapping of protected agriculture (PA) by OPAC (i.e., merging "Glasshouse", "Mulch_white" and "Tunnel") and plastic greenhouses by APGI using a Sentinel-2 product-specific threshold (reported in Table 6). "no_PA" indicates that OPAC does not predict PA and APGI does not map greenhouses, "PA" refers to OPAC predicting PA and APGI mapping greenhouses, "PA_OPAC" shows when OPAC predicts PA but OPAC does not map greenhouses and "PA_APGI" vice versa. Right panels: Google Earth Pro TCI at <3 m pixel resolution. We showed TCI of Sentinel-2 and Google Earth Pro because the latter were taken in a different date than the former.

Table 6
For the Sentinel-2 products used in the generalization assessment, we report the calculated APGI threshold, which satisfies the condition that APGI maps an equivalent extent of plastic greenhouses as OPAC did for protected agriculture, and the corresponding percentage of pixels classified as plastic greenhouses by APGI over those mapped as protected agriculture by OPAC.

Product (UTM grid square and date)	Calculated APGI threshold	Pixels classified as plastic greenhouses by APGI over those mapped as protected agriculture by OPAC (%)
T30SWF – 20210828	0.28	93.7
T31UET – 20210907	0.31	73.9
T33SVA – 20210716	0.46	63.1
T35SLU – 20210731	0.56	65.9
T35SQA - 20210827	0.52	74.6

aerial ortho-photos are actually shade nets and insect nets. When no recent images are available, site inspections (at a good timing) would be relevant to understand whether the lower values for Band 1 are due to mixed responses affected by adjacent different land covers or to the plastic material covering the vegetation. Yet, in order to avoid frequent and widespread site visits, it is crucial to establish cooperative frameworks with "entrepreneurial" farmers for learning on the management of the structures used in protected agriculture and on the observed spectral signature by Sentinel-2 or similar missions, such as Landsat 8–9.

5.2. *Classification constrained on agricultural lands to limit misclassification issues*

OPAC is constrained on cloud-free agricultural lands only. This approach is advantageous because it allows for narrowing down the classification task on a smaller area, thus being less time-demanding and less computationally-intensive. The approach is also found to be a necessity considering that clouds, potentially present anywhere, were mapped by both OPAC and APGI as protected agriculture given their high reflectance across bands. Fresh snow is not really a problem for OPAC given its high reflectance for Bands 1–8 contrasted by the low reflectance for Bands 11–12; though, patches of land partially covered with snow and vegetation can raise misclassification issues due to the mix of biomass on a highly reflective ground. We find that available cloud and snow masks implemented in the SNAP toolbox misclassify protected agriculture. Therefore, clouds and patches of snow from Sentinel-2 products in the south of Switzerland are being manually masked out in order to complete the Swiss map of protected agriculture. Brown et al. (2022) implement a workflow exploiting a new cloud mask and a time series analysis to produce the "Dynamic World" product in the Google Earth Engine. This product has a class for "Snow and ice" but not for "Clouds". It does not have a class for "Protected agriculture" neither and the pixels falling in this category are classified as "Built" or "Bare". Scientific exchanges with the authors will reveal which step of the workflow led to this classification result, thus identifying the way forward for reducing current misclassification issues due to clouds. Bare soils with high reflectance for Bands 2–4 are also challenging for OPAC. In Switzerland, these cases (based on 10 cm resolution aerial ortho-photos) refer to excavation sites, parcels close to alpine streams covered with fine alluvial sediments or very dry agricultural soils. Lithologic mapping is supported by measurements in the Very Near Infra-Red, Short Wave Infra-Red and Thermal Infra-Red wavelength regions remotely sensed using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (Rowan and Mars, 2003). It might be that from 2030 three Landsat Next satellites will deliver freely available multispectral data covering the wavelengths of Sentinel-2 and ASTER. Tests would be needed to understand whether this upcoming possibility could improve the separability of highly reflective bare soils from protected agriculture covers. Outside the proper agricultural land, the presence or not of vegetation in pixels imaging very reflective materials such as bright roads and roofs among others, again lead to misclassification issues. This aspect leads to the advantage of having a detailed mapping of agricultural lands (LWB compared to Corine).

Detailed reporting has advantages in the sense that it provides already all the necessary georeferenced information independently from remotely sensed data and detrimental weather conditions. One can also use this yearly (at best) information to avoid the computationally intensive task of image segmentation. However, detailed reporting can be affected by few limitations. For example, rather than crop rotations, only one crop is typically reported per year. Moreover, farmers may grow multiple crops with different seasonality and spectral signatures within their parcels. We face these limitations in our research and similar experiences are shared by Croci et al. (2022). More importantly, yearly reporting systems are not suitable for frequent near real-time land cover mapping, with the latter being enormously beneficial for environmental monitoring and modelling. There, Sentinel-2 finds its advantage and it is possible that a new generation of sensors with higher spatial resolution or measuring at optimal wavelengths for plastic mapping will allow for automated detection of protected agriculture. This latter data layer can be exploited for pinpointing the sources of agricultural plastic as well as improve the conceptualization of transport processes of agricultural contaminants at the catchment scale (la Cecilia et al., 2021).

5.3. Dataset development and pixel-based random forest model choice

We put effort in developing a comprehensive dataset, for a supervised classifier (i.e., random forest), that is class balanced, representative of the target classes and large enough for the model dimension (Belgiu and Drăguţ, 2016). In Section 5.1, we already discussed the important role of Band 1 and spatial accuracy for plastic greenhouse mapping. In addition, the dataset spans different locations over four seasons for two years, which allowed us to capture spatial and temporal changes of protected agriculture multispectral signatures. For example, in summer protected agriculture boosts crops yield at their maximum in warm climates (northern Europe), where the mix between high reflectance values in the RGB colors (Band 4, 3 and 2 of Sentinel-2) and the clear vegetation signature result in highly certain classification outcomes. In contrast, in hot climates (Mediterranean areas) farmers may have to white-paint the greenhouse roof or leave the soil rests, thus losing the vegetation spectra and possibly creating misclassification issues with synthetic materials or highly reflective bare soils. In winter, protected agriculture cannot always be economically sustainable in cool climates (northern Europe) due to the need of electricity to lighting and heating the greenhouse. The opposite is true in warm climates (Mediterranean areas) where the greenhouse suffice to create a favorable environment for crops growth. The sampling design influences the random forest model, and therefore, it is best practice to develop the model on comprehensive datasets (Belgiu and Drăguţ, 2016).

A random forest model is easy to develop, is ideal for dealing with the nonlinearity of variables and manages multisource and high-dimensional data (Belgiu and Drăguţ, 2016). A random forest model requires the optimization of few comprehendible parameters as compared to other machine-learning approaches, such as support vector machine (Mountrakis et al., 2011; Pal, 2005). Importantly, random forest generally achieves comparable or higher overall classification accuracies than such state-of-the-art approaches (few summarized in Table 1). The capability to deal with nonlinearity of variables is advantageous as compared to spectral index-based classification methods. These methods rely on selected threshold values, which are not homogeneous in space and time, as shown in Table 6 for the APGI. The compatibility with multisource and high-dimensional data allows for exploiting additional relevant information and state-of-the-art sensors for plastic greenhouse mapping. The acquisition of ASTRIS data fused with Sentinel-2 data may highlight the capability to better distinguish bare lands from protected agriculture. This is a valid test given that the future Landsat Next will integrate the two sources of information in a single measurement. The availability of synchronous measurements might return more accurate land cover classes. Indeed, agricultural land management and meteorological conditions can change the land cover spectra in a matter of hours (rowing, irrigating, harvesting, light-shading, etc).

The high-quality dataset and the advantages of random forest resulted in accurate classified maps in our study. The maps are achieved at high speed considering our choice to develop a pixel-based analysis. These accuracies are comparable with the ones attainable including the time- and computationally-consuming step of object-based image analysis. The latter analysis requires expert knowledge and very high-resolution images to retrieve textural properties of objects that can be used as model features. An object-based image analysis of our research sites is out of scope given that we achieved high accuracies already. Thus, we cannot compare the two approaches here. However, recent studies report misclassification of plastic greenhouses with artificial surfaces in heterogeneous region even after the application of an object-based image analysis (Balcik et al., 2020; Novelli et al., 2016). Gislason et al. (2006) reveal the importance of elevation for land cover classification of natural areas among other topographic variables including slope and aspect. However, the use of terrain features for mapping protected agriculture may limit the widespread applicability of the model. Inherently, protected agriculture alters the local, potentially unfit, environment to benefit plant growth. Yet, a training dataset with site-specific topographic patterns would affect the model output accuracy upon generalization. Our research sites involve very different environmental backgrounds and arrangement of the agricultural landscape, which lead to complex scenes aggregated in often hard to understand images (Marceau et al., 1994). Specifically to the limitations of OPAC encountered outside of agricultural areas, not even an object-based image analysis could avoid the misclassification with roads and bare lands in SPOT-7 images and some bare land surfaces in Sentinel-2 (Balcik et al., 2020).

6. Conclusions

Land cover classification is relevant for driving efficient and effective environmental monitoring as well as sustainable policies. Here, we train and validate a random forest-based land cover classifier, OPAC, for mapping open field as well as protected agriculture across different geographical landscapes and throughout two years in order to include spatial heterogeneities and seasonality patterns.

OPAC can supersede other classifiers of plastic greenhouses claimed to be generalizable at continental scales thank to its simplicity, fastness, multi-label land cover classification capability (overall accuracy of 98% and Intersection Over Union of 0.98 for predicting protected agriculture considered as a unique class in the Swiss agricultural landscape during cross-validation of the labelled dataset). Our qualitative tests in other countries show a positive outlook for mapping protected agriculture where the agricultural landscape is spatially defined (i.e, in Europe thank to Corine). We anticipate that the main issues for continental scale mapping of protected agriculture regard the availability of open access cloud-free Sentinel-2 data as input and the availability of validation points. To promote the improvement of OPAC as well as a fast mapping of protected agriculture, we make our innovative dataset (complete with annotations) open access so that users can (i) edit it, (ii) expand it and (iii) train any land cover classification model, which exploits the wavelengths measured by Sentinel-2, in any programming language.

The robust classification of protected agriculture using freely available satellite imagery remains a daunting task, subject to misclassification issues, which are mostly due to the lack of a better spatial resolution in the spectrum of wavelengths smaller than 490 nm. In order to advance this application of remote sensing further, we foresee to follow two directions. First, we plan to acquire very high resolution aerial hyperspectral data over protected agriculture to elucidate its spectral signature so as to recommend the most suitable sensors for mapping greenhouses at wavelengths down to 400 nm. Second, we realized how valuable are very high-resolution RGB imagery to gain more insights on the studied areas; however, users would face the management of high volumes of data possibly with different acquisition dates from Sentinel-2. Therefore, even more beneficial and practical, we envisage establish a collaborative framework for sharing ground truth data across research institutions and relevant stakeholders. Collaborations will allow for coping with the dynamicity of the agricultural landscape and for compiling validation points in a structured manner.

Data linking

The training set used to develop OPAC and the classification maps used in this research as well as other supporting material will be available through the Eawag data repository ERIC at https://opendata.eawag.ch/group under the project folder "NAWA-Flowpath" to ensure that these data are FAIR: Findable, Accessible, Interoperable and Reproducible.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge the Swiss Federal Office of Topography (for the very high resolution aerial ortho-photos), the Cantons (for LWB), the European Space Agency (for Sentinel-2), the Copernicus Land Service (for Corine) and Google for making their datasets freely available. DLC was supported by an Eawag Academic Transition Grant.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ophoto.2023.100033.

References

Acharki, S., Kozhikkodan Veettil, B., 2022. Mapping plastic-covered greenhouse farming areas using high-resolution PlanetScope and RapidEye imagery: studies from Loukkos perimeter (Morocco) and Dalat City (Vietnam). Environ. Sci. Pollut. Control Ser. https://doi.org/10.1007/s11356-022-23808-w.

Aguilar, M., Bianconi, F., Aguilar, F., Fernández, I., 2014. Object-based greenhouse classification from GeoEye-1 and WorldView-2 stereo imagery. Rem. Sens. 6 (5), 3554–3582. https://doi.org/10.3390/rs6053554.

Aguilar, M.A., Nemmaoui, A., Novelli, A., Aguilar, F.J., Lorca, A.G., 2016. Object-based greenhouse mapping using very high resolution satellite data and Landsat 8 time series. Rem. Sens. 8, 10–12. https://doi.org/10.3390/rs8060513.

Aguilar, M.Á., Jiménez-Lao, R., Nemmaoui, A., Aguilar, F.J., Koc-San, D., Tarantino, E., Chourak, M., 2020. Evaluation of the consistency of simultaneously acquired sentinel-2 and Landsat 8 imagery on plastic covered greenhouses. Rem. Sens. 12 (12) https://doi.org/10.3390/rs12122015.

Balcik, F.B., Senel, G., Goksel, C., 2020. Object-based classification of greenhouses using sentinel-2 MSI and SPOT-7 images: a case study from Anamur (Mersin), Turkey. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 13, 2769.

Bégué, A., Leroux, L., Soumaré, M., Faure, J.-F., Diouf, A.A., Augusseau, X., Tonneau, J.-P., 2020. Remote sensing products and services in support of agricultural public policies in Africa: overview and challenges. Front. Sustain. Food Syst. 4 https://doi.org/10.3389/fsufs.2020.00058.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogrammetry Remote Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Briassoulis, D., Dougka, G., Dimakogianni, D., Vayas, I., 2016. Analysis of the collapse of a greenhouse with vaulted roof. Biosyst. Eng. 151, 495–509. https://doi.org/10.1016/j.biosystemseng.2016.10.018.

Brown, C.F., Brumby, S.P., Guzder-Williams, B., et al., 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. Sci. Data 9, 251. https://doi.org/10.1038/s41597-022-01307-4.

Croci, M., Impollonia, G., Blandinières, H., Colauzzi, M., Amaducci, S., 2022. Impact of training set size and lead time on early tomato crop mapping accuracy. Rem. Sens. 14 (18) https://doi.org/10.3390/rs14184540.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Koetz, B., 2019. Near real-time agriculture monitoring at national scale at parcel resolution: performance assessment of the Sen2-Agri automated system in various cropping systems around the world. Rem. Sens. Environ. 221, 551–568. https://doi.org/10.1016/j.rse.2018.11.007.

Espí, E., Salmerón, A., Fontecha, A., García, Y., Real, A.I., 2006. Plastic films for agricultural applications. J. Plastic Film Sheeting 22 (2), 85–102. https://doi.org/10.1177/8756087906064220.

Fabre, C., Chow, R., Scheidegger, R., Doppler, T., Dietzel, A., Fenicia, F., Stamm, C., 2021. Can mitigation schemes produce detectable long-term temporal trends and spatial patterns in aquatic pesticide pollution? EGU Gen. Assemb. 2021 online 15263, 1.

Gao, B.-C., 1996. NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. Rem. Sens. Environ. 58, 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3.

Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random Forests for land cover classification. Pattern Recogn. Lett. 27, 294–300. https://doi.org/10.1016/j.patrec.2005.08.011.

González-Yebra, Ó., Aguilar, M.A., Nemmaoui, A., Aguilar, F.J., Urbano, D.S., 2018. Methodological proposal to assess plastic greenhouses land cover change from the combination of archival aerial orthoimages and Landsat data. Biosyst. Eng. 175, 36–51. https://doi.org/10.1016/j.biosystemseng.2018.08.009.

Harms, I.K., Diekotter, T., Troegel, S., Lenz, M., 2021. Amount, distribution and composition of large microplastics in typical agricultural soils in Northern Germany. Sci. Total Environ. 758, 143615 https://doi.org/10.1016/j.scitotenv.2020.143615.

Hasituya, Chen, Z., Li, F., Hongmei, 2017. Mapping plastic-Mulched farmland with C-band full polarization SAR remote sensing data. Rem. Sens. 9 (12) https://doi.org/10.3390/rs9121264.

Huang, Y., Liu, Q., Jia, W., Yan, C., Wang, J., 2020. Agricultural plastic mulching as a source of microplastics in the terrestrial environment. Environ. Pollut. 260, 114096 https://doi.org/10.1016/j.envpol.2020.114096.

Ji, L., Zhang, L., Shen, Y., Li, X., Liu, W., Chai, Q., Chen, D., 2019. Object-based mapping of plastic greenhouses with scattered distribution in complex land cover using Landsat 8 OLI images: a case study in Xuzhou, China. J. Indian Soc. Rem. Sens. 48 (2), 287–303. https://doi.org/10.1007/s12524-019-01081-8.

la Cecilia, D., Dax, A., Ehmann, H., Koster, M., Singer, H., Stamm, C., 2021. Continuous high-frequency pesticide monitoring to observe the unexpected and the overlooked. Water Res. X 13, 100125. https://doi.org/10.1016/j.wroa.2021.100125.

Lanz, P., Marino, A., Brinkhoff, T., Köster, F., Möller, M., 2020. The InflateSAR campaign: evaluating SAR identification capabilities of distressed refugee boats. Rem. Sens. 12 (21) https://doi.org/10.3390/rs12213516.

Lian, Z., Gu, X., Liu, L., Zhao, X., 2022. Diffuse phthalate acid esters losses induced from large amount of agricultural plastic film residues caused low risks for water quality in China during 1991-2017. J. Hazard Mater. 431, 128644 https://doi.org/10.1016/j.jhazmat.2022.128644.

Lu, L., Tao, Y., Di, L., 2018. Object-based plastic-Mulched landcover extraction using integrated sentinel-1 and sentinel-2 data. Rem. Sens. 10 (11) https://doi.org/10.3390/rs10111820.

Ma, A., Chen, D., Zhong, Y., Zheng, Z., Zhang, L., 2021. National-scale greenhouse mapping for high spatial resolution remote sensing imagery using a dense object dual-task deep learning framework: a case study of China. ISPRS J. Photogrammetry Remote Sens. 181, 279–294. https://doi.org/10.1016/j.isprsjprs.2021.08.024.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. In: Sen2Cor for Sentinel-2, vol. 10427. SPIE.

Marceau, D.J., Howarth, P.J., Gratton, D.J., 1994. Remote sensing and the measurement of geographical entities in a forested environment. 1. The scale and spatial aggregation problem. Rem. Sens. Environ. 49, 93–104. https://doi.org/10.1016/0034-4257(94)90046-9.

Mormile, P., Stahl, N., Malinconico, M., 2017. The world of plasticulture. In: Soil Degradable Bioplastics for a Sustainable Modern Agriculture. Springer, p. 1.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogrammetry Remote Sens. 66, 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001.

Nemmaoui, A., Aguilar, M.A., Aguilar, F.J., Novelli, A., García Lorca, A., 2018. Greenhouse crop identification from multi-temporal multi-sensor satellite imagery using object-based approach: a case study from Almería (Spain). Rem. Sens. 10 (11) https://doi.org/10.3390/rs10111751.

Novelli, A., Aguilar, M.A., Nemmaoui, A., Aguilar, F.J., Tarantino, E., 2016. Performance evaluation of object based greenhouse detection from Sentinel-2 MSI and Landsat 8 OLI data: a case study from Almería (Spain). Int. J. Appl. Earth Obs. Geoinf. 52, 403–411. https://doi.org/10.1016/j.jag.2016.07.011.

Ntinas, G.K., Neumair, M., Tsadilas, C.D., Meyer, J., 2017. Carbon footprint and cumulative energy demand of greenhouse and open-field tomato cultivation systems under Southern and Central European climatic conditions. J. Clean. Prod. 142, 3617–3626. https://doi.org/10.1016/j.jclepro.2016.10.106.

Odermatt, D., Gege, P., 2022. Lake colors: interpreting apparent optical properties. In: Encyclopedia of Inland Waters, pp. 474–489.

Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Rem. Sens. 26, 217–222. https://doi.org/10.1080/01431160412331269698.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the great plains with ERTS. In: Freden, S.C., Mercanti, E.P., Becker, M. (Eds.), Third Earth Resources Technology Satellite–1 symposium.Volume I: Technical Presentations, NASA SP-351. NASA, Washington, D.C., pp. 309–317

Rowan, L.C., Mars, J.C., 2003. Lithologic mapping in the mountain pass, California area using advanced Spaceborne thermal emission and reflection radiometer (ASTER) data. Rem. Sens. Environ. 84, 350–366.

Schwanninger, M., Rodrigues, J.C., Fackler, K., 2011. A review of band assignments in near infrared spectra of wood and wood components. J. Near Infrared Spectrosc. 19, 287–308. https://doi.org/10.1255/jnirs.955.

Sun, H., Wang, L., Lin, R., Zhang, Z., Zhang, B., 2021. Mapping plastic greenhouses with two-temporal sentinel-2 images and 1D-CNN deep learning. Rem. Sens. 13 (14) https://doi.org/10.3390/rs13142820.

Tsendbazar, N., Herold, M., Li, L., Tarko, A., de Bruin, S., Masiliunas, D., Duerauer, M., 2021. Towards operational validation of annual global land cover maps. Rem. Sens. Environ. 266 https://doi.org/10.1016/j.rse.2021.112686.

UST, 2021. Agricoltura e alimentazione, Statistica tascabile 2021. Ufficio federale di statistica, pp. 1–24.

van Delden, S.H., SharathKumar, M., Butturini, M., Graamans, L.J.A., Heuvelink, E., Kacira, M., Marcelis, L.F.M., 2021. Current status and future challenges in implementing and upscaling vertical farming systems. Nat. Food 2, 944–956. https://doi.org/10.1038/s43016-021-00402-w.

van der Velden, N.J.A., Janse, J., Kaarsemaker, R.C., Maaswinkel, R.H.M., 2004. Suatainability of greenhouse fruit vegetables: Spain and The Netherlands; development of a monitoring system. ISHS Acta Hortic. 655, 275–281. https://doi.org/10.17660/ActaHortic.2004.655.34.

Wang, J., Li, J., Liu, S., Li, H., Chen, X., Peng, C., Liu, X., 2021. Distinct microplastic distributions in soils of different land-use types: a case study of Chinese farmlands. Environ. Pollut. 269, 116199 https://doi.org/10.1016/j.envpol.2020.116199.

Yang, D., Chen, J., Zhou, Y., Chen, Xiang, Chen, Xuehong, Cao, X., 2017. Mapping plastic greenhouse with medium spatial resolution satellite data: development of a new spectral index. ISPRS J. Photogrammetry Remote Sens. 128, 47–60. https://doi.org/10.1016/j.isprsjprs.2017.03.002.

Zhang, P., Du, P., Guo, S., Zhang, W., Tang, P., Chen, J., Zheng, H., 2022. A novel index for robust and large-scale mapping of plastic greenhouse from Sentinel-2 images. Rem. Sens. Environ. 276 https://doi.org/10.1016/j.rse.2022.113042.

Zhou, Z., Kaufmann, H., Bohn, N., Bochow, M., Kuester, T., Segl, K., 2022. Identifying distinct plastics in hyperspectral experimental lab-, aircraft-, and satellite data using machine/deep learning methods trained with synthetically mixed spectral data. Remote Sens. Environ. 281, 113263 https://doi.org/10.1016/j.rse.2022.113263.