Cover page

Exam information
NIGK18800E - Environmental Science Thesis 30 ECTS, Department of Geosciences and Natural Resource Management - Kontrakt:119755 (Marcel Mohr)

Handed in by
Marcel Mohr
mpw402@alumni.ku.dk

Exam administrators
Eksamensteam, tel 35 33 64 57
eksamen@science.ku.dk

Assessors
Alexander Prishchepov
Examiner
alpr@ign.ku.dk
+4535331386

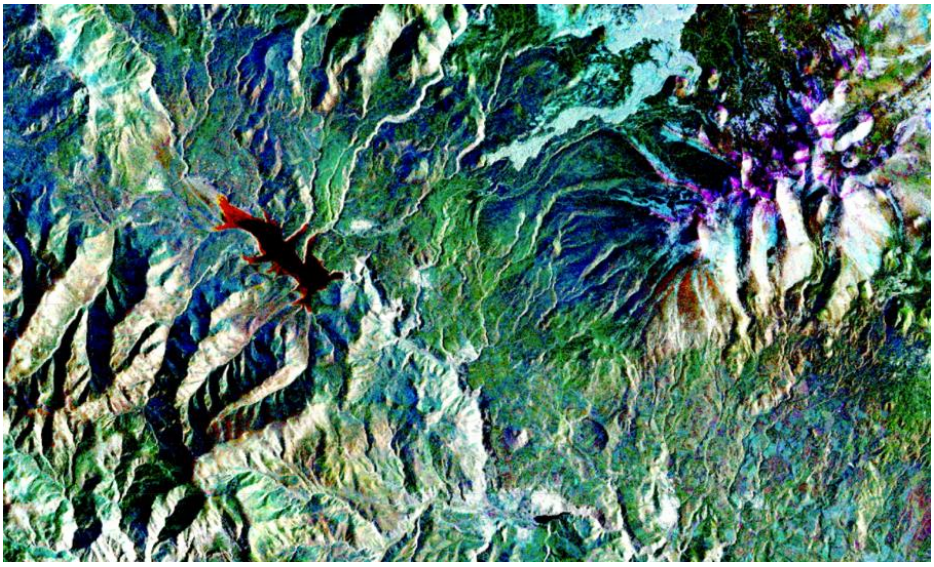
Hand-in information

Titel: Towards mapping the nature value of agricultural land use in the Armenian Highlands
Titel, engelsk: Towards mapping the nature value of agricultural land use in the Armenian Highlands
Tro og love-erklæring: Yes
Indeholder besvarelsen fortroligt materiale: No

UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

Master of Science Thesis

Marcel Mohr



Towards mapping the nature value of agricultural land use in the Armenian Highlands

Potentials and limitations of classifying smallholder farming in complex landscapes using Google Earth Engine

UNIVERSITY OF HOHENHEIM

AUA
American University of Armenia

Department	Department of Geosciences and Natural Resource Management Section for Geography
Author	Marcel Mohr (MPW402)
Study Program	Master of Science “Environmental Science – Soil, Water, Biodiversity” Universities of Hohenheim & Copenhagen
Supervisor	Professor Alexander Prishchepov, University of Copenhagen
Co-supervisor	Professor Claudia Bieling, University of Hohenheim
Submitted	09.02.2021

Keywords	Remote sensing, LULC classification, smallholder farming, Armenian Highlands, nature value, random forest, Sentinel-2, Sentinel-1, topography metrics, nighttime lights, texture, Google Earth Engine

Acknowledgements

First and foremost, I want to express my deep gratitude to my supervisor Alexander Prishchepov for his extraordinary advice, support and mentoring over the course of this project. I likewise want to extend my gratitude to my co-supervisor Claudia Bieling for her support and for enabling the collaboration with GAtES and the American University of Armenia. Alen Amirkhanian and Gohar Shahinyan have been supportive as contact persons in Armenia and helped with translation problems. Thanks also goes to Planet Labs for providing access to useful satellite imagery. My thanks extend to everyone else that have shared their knowledge and kindness with me over the last months and made this project possible. I also want to show my gratitude to my partner and friend Elinor Boos for all the love and support. Finally, I want to thank my parents for their support that has allowed me to choose and shape my path freely for all these years.

List of Abbreviations

BSI	Bare soil index
CORINE	Coordination of information on the environment
EU	European Union
ESA	European Space Agency
GAtES	German-Armenian Network on the Advancement of Public Participation GIS for Ecosystem Services as a Means for Biodiversity Conservation and Sustainable Development
GDP	Gross domestic product
GEE	Google Earth Engine
GLCM	Grey level co-occurrence matrix
GRD	Ground range detected
LULC	Land use and land cover
MMU	Minimum mapping unit
MSI	Multi-spectral imaging instrument
NASA	National Aeronautics and Space Administration
NIR	Near infrared
NDI	Normalized difference index
RF	Random forest
RGB	red, green, blue (visible spectra)
ROI	Region of interest
SAR	Synthetic aperture radar
SRTM	Shuttle Radar Topography Mission
SWIR	Short wave infrared
VIIRS	Visible Infrared Imaging Radiometer Suite
VHSR	Very high spatial resolution

List of figures

List of tables

Abstract

The highlands of Armenia are a prime example for mountainous landscapes featuring smallholder-based, dispersed and multifunctional land use. These areas often have high nature value and provide various ecosystem services. Yet, detailed and accurate mapping efforts with satellite imagery are facing difficulties due to the spatial and spectral complexity of intercepted landscapes. This study aims at exploring the potentials and limitations of mapping smallholder farming with an increased thematic detail in a 4726 km^2 study area in Armenia using Google Earth Engine (GEE). For this purpose, a coherent classification catalogue was developed featuring 11 land use and land cover (LULC) classes in total and six agricultural target classes. The proposed classification system follows a classic pixel-based approach. The input data comprises intra-annual and multi-temporal optical imagery as well as other multisource data, all available within GEE. The data was processed to construct composites with filtered image spectra and fixed temporal bins. Different input datasets were created to assess the added value for map predictions through the multisource data. The input datasets were classified with random forest and acquired training data. For classification the study area was stratified due to GEE quota limitations. The map predictions were assessed through an accuracy assessment supported by a comprehensive validation sample. Based on the results, a "good choice" dataset was constructed that achieved overall accuracies of 76% and 82% for the two sub-regions. User's accuracies ranging from 57% to 88% indicate mapping constraints of the highly collinear agricultural target classes. Compared to a baseline dataset with optical data only, an added value of the multisource data could be observed. The improvements are however limited, suggesting other accuracy limitations such as spatial and temporal resolution. In respect to this, the mapping approach can be further improved without sacrificing its simplicity and practicality. The approach shows potential regarding the transferability but also faces limitations regarding the workload for training data. The produced map delivers accurate and detailed information of the extent and the distribution of the LULC classes. Such data is of great relevance for many current land-use challenges in Armenia.

Table of Contents

1. Introduction

Agricultural land use in mountainous regions is often smallholder based and rather complex regarding the spatial structure resulting in limitations for mapping efforts. Most progress in land use and land cover (LULC) mapping has been made so far for large scale agriculture areas. It remains challenging to accurately map small scale, dispersed and multifunctional land use in intercepted landscapes. Armenia is a prime example of such mountainous regions with complex agricultural and semi-natural landscapes. The small and landlocked country in the Caucasus region exhibits a rich biodiversity and a long history of agricultural land use (CBD, 2014). Agriculture in a mountainous environment is often restricted due to climatic or topographic factors. These limitations are usually overcome by the local population through intelligent and sustainable adaptations of the agricultural production systems. Agriculture in Armenia and other mountainous regions is thus traditionally characterized by extensive management systems within a subsistence-based family farm setting (Wymann von Dach *et al.*, 2013; FAO, 2020b). The resulting closely intertwined and dynamic social-ecological systems shaped the mountainous agricultural landscapes for centuries (Plieninger and Bieling, 2012).

1.1. Nature value of mountainous farming landscapes

Contrasting to the inherent low productivity, mountainous farming regions often exhibit a high nature value and feature attractive cultural landscapes. Besides the primary functions of producing food and fibre, these areas are usually rich in natural or semi-natural vegetation and are able to support species and habitats of conservation value which specifically depend on the low-intensity land use (Lomba *et al.*, 2020). Within the associated social-ecological systems a broad range of ecosystem services can be provided for the human well-being. Supporting and regulating services include nutrient cycling, primary production, soil formation, maintenance of genetic resources, pollination, biological control and storage and purification of water as well as climate, soil erosion and flood regulation. Cultural services such as recreation, sense of place and maintenance of cultural heritage and scenic landscapes are often undervalued but particularly important for Armenia and many mountainous regions. Mountainous farming landscapes experience a recent trend of reduced resilience due to different socio-economic and institutional changes such as rural depopulation, termination of subsidies or evolving land-use conflicts. This

often results in abandonment or afforestation of marginal farming areas and intensification of the most productive areas causing a widespread decline of extensive management systems (Prishchepov *et al.*, 2013; Wymann von Dach *et al.*, 2013; Lomba *et al.*, 2015; Buchner *et al.*, 2020).

The nature value of farmland is a concept adopted in the Common Agricultural Policy, the EU legislation on agriculture. Based on a conceptualization of the nature value of farmland the extent of farmland with a high nature value is monitored in a spatial explicit way using different data sources including CORINE land cover data as well as farming and environmental data (Lomba *et al.*, 2014). While the assessment and monitoring of the nature value of farmland in the EU is far from perfect, such instruments are still largely absent in many parts of the world including Armenia. The assessment of the nature value of farmland in Armenia may however enable to characterize areas with conservation value and direct financial support to those agriculture-dominated landscapes where high nature value is given and dependent on the continuation of low-intensity farming systems (Lomba *et al.*, 2015).

1.2. Post-soviet smallholder agriculture in Armenia

There is comprehensive evidence that the Armenian Highlands have an agricultural history dating back thousands of years (Avetisyan, 2010; Wilkinson *et al.*, 2012). As a former Soviet country, agriculture in Armenia experienced drastic changes during the recent past. The establishment of the Soviet power in 1920 resulted in profound structural changes of the traditional subsistence-based agriculture. As in the rest of the Soviet Union, farm households were deprived of private property, forced to collectivize and state ownership was established. Consequently, large scale irrigation, land improvement and orchard planting programs where implemented which allowed the food and agriculture sector to become one of the most important pillars of Armenia's economy (Avetisyan, 2010).

After the Armenian independence in 1991, the role of the agricultural sector within the country's economy increased to more than 30% of the gross domestic product (GDP) and more than 40% of total employment compared to a contribution of less than 20% of each prior to independence. This development was linked to strong individualized land rights and dramatical shift to individual farming to ensure food security needs of the population confronted with

uncertainties during the first transition phase towards an independent economy as well as the associated collapse of non-agricultural sectors in the early 1990s (Rozelle and Swinnen, 2004; Millns, 2013). With the stabilisation of the economy, the agricultural share of the GDP dropped again resulting in a contribution of around 18% of the country's GDP and 35% of total workforce (state: 2018) (World Food Programme, 2019). The 2008 global financial crisis heavily impacted the experienced rapid economic growth and poverty mitigation of the early post-independence period. As a result, rural poverty has remained high, ranging around 35% of the population in 2012 and 29% in 2019. Poverty and food insecurity are particularly an issue for the rural population. Women and the youth are also overrepresented amongst the poor (Millns, 2013; World Food Programme, 2019).

Today, the agricultural sector in Armenia is predominantly smallholder-based, consisting of 360,000 held properties with an average size of 1.5 ha (World Food Programme, 2019). 42% of the smallholder farms are smaller than 0.5 ha (FAO, 2020b). 88% of the farms are smaller than 2 ha and cover 77% of the total land area. The farm sizes are varying between 0.61 ha in the Ararat valley to 3 ha in the uplands (Millns, 2013). On average, each farmer owns three parcels of land, of which one is irrigated and two are non-irrigated. Further, 15% of the farmers cultivate land that has been leased with an average size of 3.2 ha. This fragmented and subsistence-based profile of the sector is an immediate result of the disintegration of the Soviet Union and the associated privatization. During the early 1990s, 869 large collective and State farms with 147,000 separate parcels were privatized to create 338,000 farms and rural households with more than 1.2 million relatively small plots of land at an unprecedented speed (Avetisyan, 2010; FAO, 2020b).

The promptness and uncoordinated character of this process led to partial resentments and conflicts arising over the allocation of water rights and the distribution of fundamental materials and equipment (Curtis and Library of Congress, 1995; Millns, 2013). Following the privatization, livestock numbers declined, the total area of irrigated lands decreased by half and the use of fertilizers decreased by two-thirds (Worldbank, 2012). The distribution of land was rather chaotic and followed a proportional scheme where the land of collective farms was distributed among community members. Communities with higher population density received smaller parcels, resulting in highly fragmented land ownership. The land each community member received could

comprise a large number of non-contiguous parcels often with long distance from the homestead to the parcels and between the parcels (Hartvigsen, 2014; Buchner *et al.*, 2020).

Armenia is a land-scarce country, yet around 50% of both arable land and former pastures across Armenia are not cultivated or managed and one third of the farmers with agricultural holdings do not cultivate their land at all (Millns, 2013; Urutyan, 2020). Besides land fragmentation, there are several challenges smallholders in Armenia are facing today that contribute to ineffective farmland use. Such challenges are amongst others, the lack of farming and entrepreneur skills, missing access to information and markets, aging rural population and out-migration of youth, over-indebtedness, unwillingness to cooperate and limited access to mechanization and technology (Millns, 2013; FAO, 2020a; Urutyan, 2020). Over-indebtedness seems to become an increasing problem for many of the smallholders in rural areas. Urutyan (2020) emphasizes that there might be an issue regarding the inverse trends in credits to agriculture and effective agricultural land use or agricultural outputs (see figure 1). Smallholders are borrowing more and more money to pay off existing depts while abandonment of the fragmented agricultural land increases.

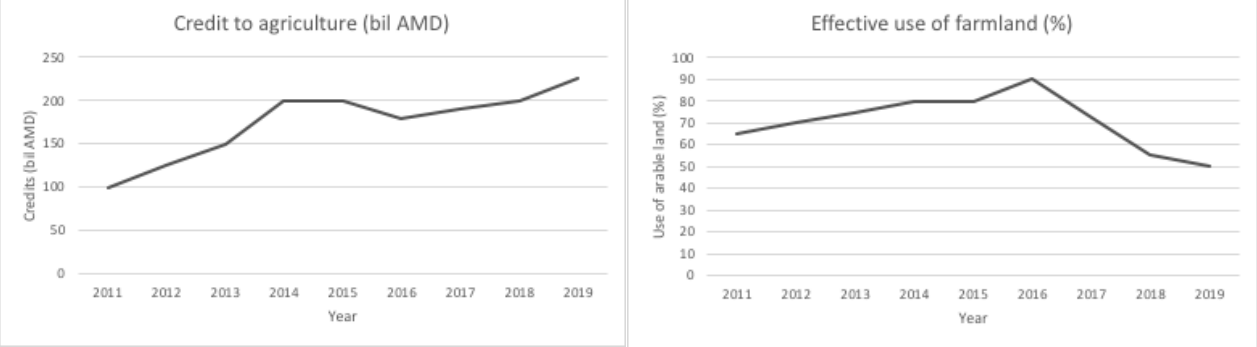


Figure 1 Credits to agriculture in billion Armenian Dram (AMD) and use of total arable land (%). Whereas the credits are constantly risings the use of arable land decreases sharply since 2016 (Urutyan, 2020).

The situation of smallholders in Armenia as "land-users lost in transition" and the associated issue of ineffective use of farmland poses risks for both mountain ecosystems, being vulnerable to changes in land use as well as rural livelihoods facing challenges regarding perpetuating poverty and unemployment. As a result, losses in biodiversity and decreased resilience of the socio-ecological systems as well as alterations in the ability of ecosystems to provide critical goods and services largely limit sustainable development of human and natural resources (Muller *et al.*, 2009; Theissen *et al.*, 2019). Livestock farming is for example crucial for the food production in

mountainous regions of the northern hemisphere. Grassland management has thus shaped the biodiversity and floristic composition in those cultural landscapes for centuries. Land-use change, due to structural changes in agriculture or socio-economic shifts, often leads to deterioration of ecosystem functioning and the resilience of these landscapes (Theissen *et al.*, 2019).

There are various efforts in Armenia towards reaching the Sustainable Development Goals including conservation of terrestrial and in particular mountainous ecosystems, sustainable agriculture, food-security and alleviation of rural poverty and climate change mitigation (Republic of Armenia, 2018). Policies and programs for afforestation, land consolidation, sustainable intensification of agriculture and wildlife conservation are however often conflicting and require land-use trade-offs (Foley *et al.*, 2005). Besides afforestation and wildlife conservation, there are priority interests in restructuring rural livelihoods and smallholder agriculture to provide circumstances that allow a sustainable intensification and gender equality among land users (Republic of Armenia, 2018; World Food Programme, 2019).

Although much is at stake, there is a lack of information and spatial data related to post-soviet smallholder agriculture in Armenia which is hindering efforts. In particular fine-scale land-use maps are often unavailable or of unknown reliability (Kuemmerle *et al.*, 2009). The data gap regarding detailed spatial-temporal information of LULC is limiting the success of endeavours towards sustainable development. Policy and decision makers are often “flying blind” without means to understand the holistic and complex character of food and agricultural production systems in mountainous landscapes (Urutyan, 2020).

1.3.Remote sensing in land system science

Satellite images and other remote sensing data are valuable means to feasibly classify, map and monitor LULC over large areas across the globe. In the field of land system science, agricultural land use and land-use change is increasingly explored with remote sensing due to innovations regarding sensors and data processing capabilities in recent years. Besides the classification of cropping systems and practices such as crop type, cropping intensity or management regimes, the change mapping of agricultural land use has been investigated intensively. In the post-soviet context, agricultural land-use dynamics have been explored across Kazakhstan (Kraemer *et al.*, 2015), Russia, Belarus, Lithuania, Latvia and Poland (Alcantara *et al.*,

2013), Slovakia and Romania (Kuemmerle *et al.*, 2009; Griffiths *et al.*, 2013) and recently also across the Caucasus (Yin *et al.*, 2018; Buchner *et al.*, 2020).

LULC maps with a high thematic detail of spectrally narrow agricultural classes are also increasingly explored as in case studies of agricultural land use in Germany (Griffiths, Nendel and Hostert, 2019), the Brazilian cerrado (Bendini *et al.*, 2019), Australia (Brinkhoff, Vardanega and Robson, 2020) or Belgium (Van Tricht *et al.*, 2018). Challenges in remote sensing regarding heterogenous landscapes and complex small-scale subsistence farming also received increased attention in recent years (Senf *et al.*, 2015; Lebourgeois *et al.*, 2017). Detailed mapping of smallholder-based farming areas remains challenging due to various factors including: high spatial collinearity amongst land-use classes, small patch sizes and fragmented distribution of the fields as well as low intensities of the land-use signals (Lebourgeois *et al.*, 2017).

As many technological fields, remote sensing experienced an accelerated development during the last decade (Splinter, Harley and Turner, 2018; Zhu *et al.*, 2018). The main quality properties of (optical) spaceborne remote sensing data improved substantially, including: (1) Spatial resolution (i.e. the size of the acquisition unit/ pixel); (2) Spectral resolution (i.e. the number of bands with certain content of information such as spectra or acquisition mode); (3) Temporal resolution (i.e. the revisit time or the time between two consecutive acquisitions of the same ground area) (Dash and Ogutu, 2015). These properties are important factors for a detailed spatial assessment of smallholder agriculture in heterogenous mountainous landscapes (Lebourgeois *et al.*, 2017).

The spatial resolution directly limits mapping efforts of small agricultural plots as the smallest minimum mapping unit needs to be at least of a similar size or smaller. Optical sensors like the Sentinel-2 multi-spectral instrument (MSI) have high potential for mapping agricultural land use in smallholder dominated landscapes (Lebourgeois *et al.*, 2017). Besides the moderate to high spatial resolution of 10–60 m, this mainly arises from the spectral and temporal resolution with 13 bands and a revisit time of five days in the constellation. Although high to very high spatial resolution sensors such as Rapid-Eye or Worldview are also recognized as useful for mapping smallholder farming, the lower spectral resolution with only three to four bands in the visible spectrum (RGB) and near infrared (NIR) can limit a detailed LULC classification and often require more complex deep learning applications due to the high spatial complexity (Du *et al.*, 2019).

The temporal resolution is of particular significance for multi-temporal analysis of vegetation phenology, which is widely used to classify different crops or agricultural management regimes (Bendini *et al.*, 2019; Rufin *et al.*, 2019). The multi-temporal images can be used to display phenological changes and differences between vegetation and crop types. Temporal gaps due to cloud cover can be overcome and the cost of computation reduced by temporal binning of intra-annual image time series, where all available observations within a defined time period (e.g. a month or ten days) are used to create quality composites (e.g. median composites). The temporal resolution of Sentinel-2 can be further improved due to the interoperability with Landsat-8 operational land imager. This allows shorter temporal bins to adequately reflect the phenological differences between LULC classes.

In comparison to single images or composites, multi-temporal and -spectral image analysis usually requires appropriate computational power given the large number of bands and image dates. This is especially the case for the classification of vegetation, cropping practices and intensities, where multiple approaches to capture the phenology using multi-temporal and -spectral data have been explored within the remote sensing community (Senf *et al.*, 2015; Griffiths, Nendel and Hostert, 2019; M. Zhang *et al.*, 2019). Spectral indices and transformations are means to condense the amount of multi-spectral data and derive indicators for different ground features. Indices for vegetation features such as the Normalized Difference Vegetation Index (NDVI) or the Tasselled Cap metrics are widely acknowledged to map LULC from national to continental scales (Teluguntla *et al.*, 2018; Rufin *et al.*, 2019). The spectral derivates are simple and effective algorithms for quantitative and qualitative evaluations and can enhance classification due to their ability to detect different land cover types (e.g. vegetation or urban structures) (Xue and Su, 2017). The mapping of permanent crops in the Maipo valley in Chile showed, that the utilization of all the reflectance bands from multi-spectral imagery improves accuracy over using a small number of normalized difference indices (NDIs). However, when combining many NDIs, best accuracy results can be achieved (Peña and Brenning, 2015; Peña, Liao and Brenning, 2017). Yet, the dimensionality of the imagery data directly affects the classification outcome and redundancies within the data should be minimized (Belgiu and Drăgu, 2016).

Several studies showed beneficial effects of combining optical imagery and imagery acquired by active synthetic aperture radar (SAR) sensors (Stefanski *et al.*, 2014; Van Tricht *et al.*, 2018; Mahdianpari *et al.*, 2019; Brinkhoff, Vardanega and Robson, 2020). Besides cropland and wetland mapping, this is in particular the case for scattered woody vegetation (Baumann *et al.*, 2018; W. Zhang *et al.*, 2019). Due to the independence from weather conditions, a combination with SAR data is especially useful in mountainous or tropical regions with consistent cloud cover over longer periods (Zhang *et al.*, 2020).

Ancillary continuous data such as topography metrics derived from digital elevation models, precipitation, soil moisture and nighttime lights are also increasingly recognized to improve LULC classifications or change detection (Lebourgeois *et al.*, 2017; Chen *et al.*, 2018; Yuan *et al.*, 2019; Gumma *et al.*, 2020). There is evidence that specifically in mountainous areas with complex terrain, the use of topographic data can enhance mapping accuracies (Lebourgeois *et al.*, 2017; Adepoju and Adelabu, 2020). Stable nighttime lights data is typically used for socio-economic activities and the detection of urban areas or rural settlements. Texture metrics quantitatively describe relationships of digital number values of neighbouring pixels and can be a mean to aid the classification of LULC. Since spatial relationships are often uncorrelated with the spectral data of a given class, the inclusion of specific texture metrics may improve classification accuracies (Hall-Beyer, 2017; Numbisi, Van Coillie and De Wulf, 2018).

These novel developments exhibit great potential for a detailed mapping of smallholder agriculture within a heterogenous mountainous landscape, but require the processing of a vast amount of multi-temporal and -spectral remote sensing data. Within recent years, the increasing availability of large-volume open access remote sensing data and the development of advanced machine learning applications have been integrated with powerful cloud computing resources such as Amazon's Web Services and Google Cloud. GEE is addressing the issues related to big data of earth observation. GEE is an open access, cloud-based platform for parallel processing of petabyte-scale data. It is based on a client-server programming model allowing batch processing using JavaScript on a dedicated application programming interface (Gorelick *et al.*, 2017; Kumar and Mutanga, 2018). GEE hosts a large pool of satellite imagery and geospatial datasets, and allows web-based algorithm development and results visualization in a reasonable processing time. Besides the computing and storage capacity, a number of well-known machine learning

algorithms have been implemented and linkages to deep learning applications such as TensorFlow have been established (Tamiminia *et al.*, 2020).

The development of advanced machine learning tools further eases the handling of large multi-temporal and -spectral remote sensing data. Traditional classifiers insufficiently manipulate complex and high-dimensional data and depend on the statistical distribution of the input data. Random Forest (RF) is a widely used advanced machine learning algorithm that can handle large amounts of uncorrelated data and deliver high classification accuracies across applications. RF is less sensitive to noise and overfitting and relatively easy to execute given the low number of parameters compared with other machine learning algorithms such as Support Vector Machine (Mahdianpari *et al.*, 2019).

1.4.Study objective, framework & research questions

This study aims to explore the potential and limitations of mapping smallholder farming with an increased thematic detail in a heterogenous mountainous landscape in Armenia. The agricultural land use with six target classes is complemented by other semi-natural or natural and artificial classes, resulting in a total of 11 LULC classes within the study region. For a detailed typology of the classes see section 2.2. The classification approach follows the classic concept of classifying all classes simultaneously as opposed to a hierarchical approach with several classification stages[1] or a sequential approach where one class at a time is classified against a grouped class. With the classic approach, limitations regarding class specific confusion and misclassification can be accessed in great detail. In contrast, hierarchical and sequential approaches are more time consuming and the assessment of specific class confusion is limited by grouped classes and the use of masks. Finally, the proposed approach does not include an "other" class, resulting in a cohesive LULC map for the entire study region. The classification approach in this study is exclusively based on open access data and open-source software for processing Including GEE, the R Project for Statistical Computing and QGIS of the Open Source Geospatial Foundation.

[1] i.e. first classifying broad classes like cropland and woody vegetation and using these produced mask layers to subsequently classify different types of cropland or woody vegetation.

Due to the SARS-CoV-2 pandemic, an exploration of the smallholder farming systems and a collection of references data in the field was not possible. The proposed classification framework is therefore based entirely on remotely acquired training and validation data. The assessment of class membership within the sampling schemes is based solely on visual interpretation of multi-temporal very high spatial resolution images (VHSR) and other available spatial-temporal data. For a detailed explanation of this assessment framework of class membership see section 2.7 and Appendices II-IV.

The LULC is classified with RF machine learning algorithm due to its practicality and high mapping accuracies. The classification approach is based on intra-annual multi-temporal remote sensing data. Multi-spectral optical data of the Sentinel-2 constellation is used as primary classification input. The reflectance bands from multi-spectral imagery are used to derive several NDIs and spectral transformations. SAR data of the Sentinel-1 constellation and ancillary data, namely topography metrics, texture metrics and nighttime lights are explored regarding the added value for the classification outcome. Based on this data pool, different datasets are constructed in order to assess the differences in the RF model accuracies of the associated datasets. This is done to evaluate the use of the multisource data for mapping smallholder agriculture within the study region in the Armenian highlands. A detailed description of this approach is provided in sections 2.3 and 2.5.

McRoberts (2011) emphasizes that maps are usually nothing more than pretty pictures without a rigorous accuracy assessment and indication of the precision of these estimates. In order to produce meaningful model predictions this study followed the good practice recommendations for accuracy and area assessment as provided by Olofsson *et al.* (2014). Accordingly, the sampling of validation data used to derive the map accuracy is implemented with statistical inference as scientific support for a generalization from the sample data to the population parameters. For a detailed description of the sampling procedure and accuracy assessment see section 2.7 and 2.8.

The main objective of this study is to estimate the extent and distribution of the LULC classes in the study region. In line with this, the mapping potential and limitations are aimed to be explored. Different remote sensing data types are assessed regarding the added value for the LULC classification. The central research question of this study thus is: “How accurate and how

detailed is the proposed mapping approach of smallholder agriculture in a complex mountainous landscape?" This includes the research sub-question: "What are the effects on map predictions when including multisource data?" Within this research framework other relevant questions include:

- What is the effect of combining optical and SAR data?
- What is the effect of using spectral derivates (NDIs/spectral transformations)?
- What is the effect of including ancillary data such as texture metrics, topography metrics or nighttime lights?

Based on the results of the accuracy assessment for the different datasets these questions are evaluated in the discussion and potentials and limitations are determined. This includes considerations regarding the depth of the class typology as well as general characteristics of the classification system such as practicality of the classification approach, sufficiency of the spatial, temporal and spectral resolution, transferability and ease of training and validation data collection. Other important points of discussion concern the reliability of LULC mapping without ground truth data and interpreter bias as well as the overall performance in the light of other similar mapping efforts.

2. Materials and methods

2.1.Study region

The study region in the south of Armenia comprises the entire administrative district of Vayots Dzor with the sub-districts Yeghegnadzor and Vayk. This area is extended by the northern part of Syunik with the sub-districts Sisian and Goris (see figure 2). In total the study region covers an area of 4,725.8 km^2. This chapter provides an overview of the physical geography and agricultural land use of the study area. This includes a description the mountain climate and vegetation zones as well as the associated biodiversity and environmental threats through agriculture. Further focus is put on the agricultural land-use systems and the smallholder livelihoods.

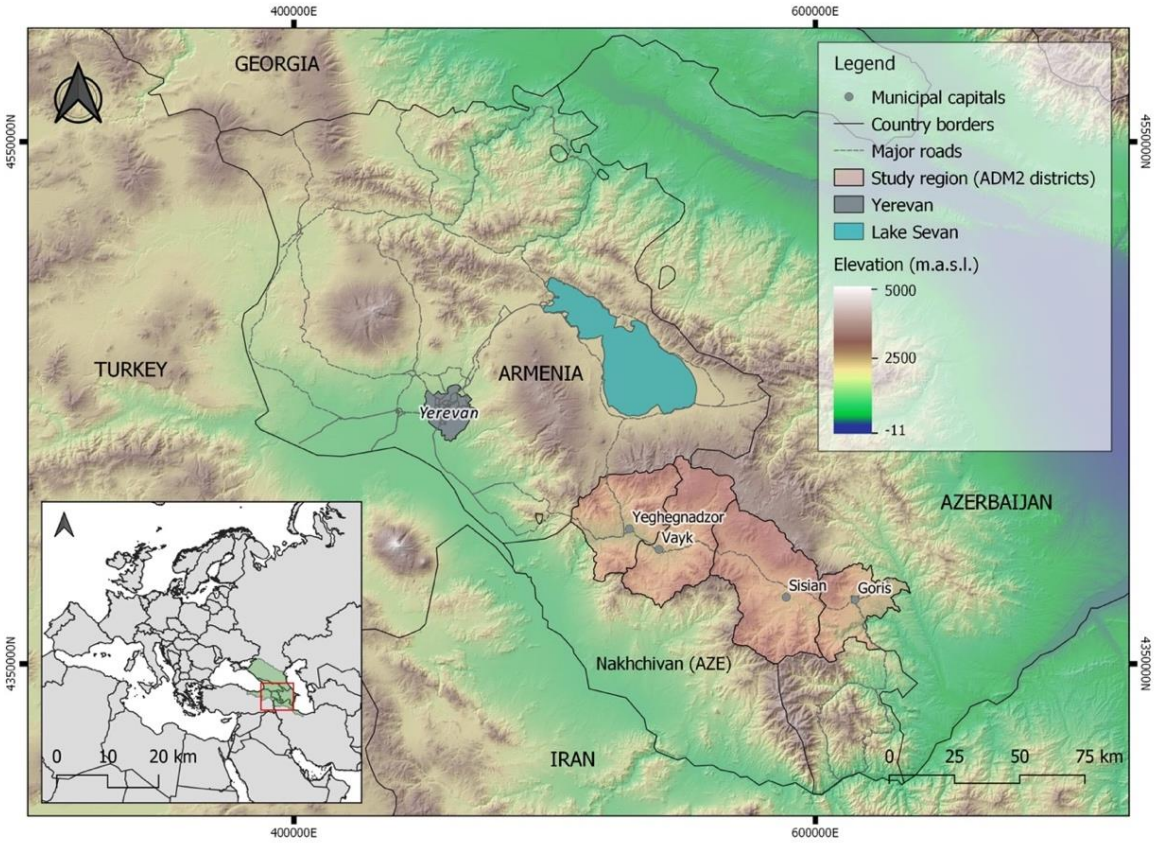


Figure 2 Overview map and location of the study region in the Armenian Highlands. The study region is depicted in light red with the four sub-regional districts (AMD2) named according to their municipal capitals. The inset map is showing the location of the Caucasus Ecoregion (green) and the map section is corresponding to the red square. As for all maps the projection is UTM WGS84 (Data: OSM Roads; Shapefiles AUA; SRTM DEM; EC World Countries; source: own illustration).

Located within the Zangezur mountain ridge of the lesser Caucasus, the region has a mean altitude of 2,125 meters above sea level (m.a.s.l). With a minimum altitude of around 700 m in the main valleys and a maximum altitude of approximately 3,500 m on the mountain peaks the landscape is structured by the elevation and comprises heterogenous orographic landscape with deep gorges and canyons, volcanic plateaus and steep mountain ridges. The topographic heterogeneity is responsible for a pronounced climatic variability resulting in a high diversity of plant communities (figure 3). Besides the orographic climatic variability, there is a main line of separation regarding the distribution of the annual precipitation. Areas of similar elevation receive double annual precipitation in Synuik (Goris, 627mm) compared to Vayots Dzor (Yeghegnadzor, 358mm) (CLIMATE-DATA.ORG, no date).

Due to the continental arid conditions and the high elevation, natural grassland ecosystems such as steppe grassland, meadow steppe as well as subalpine and alpine meadows dominate. The lower mountain belt (375-1200 m.a.s.l.) in Vayots Dzor exhibits semi-deserts with gypsophilous and halophilous plant communities which show specific adaptations to the arid

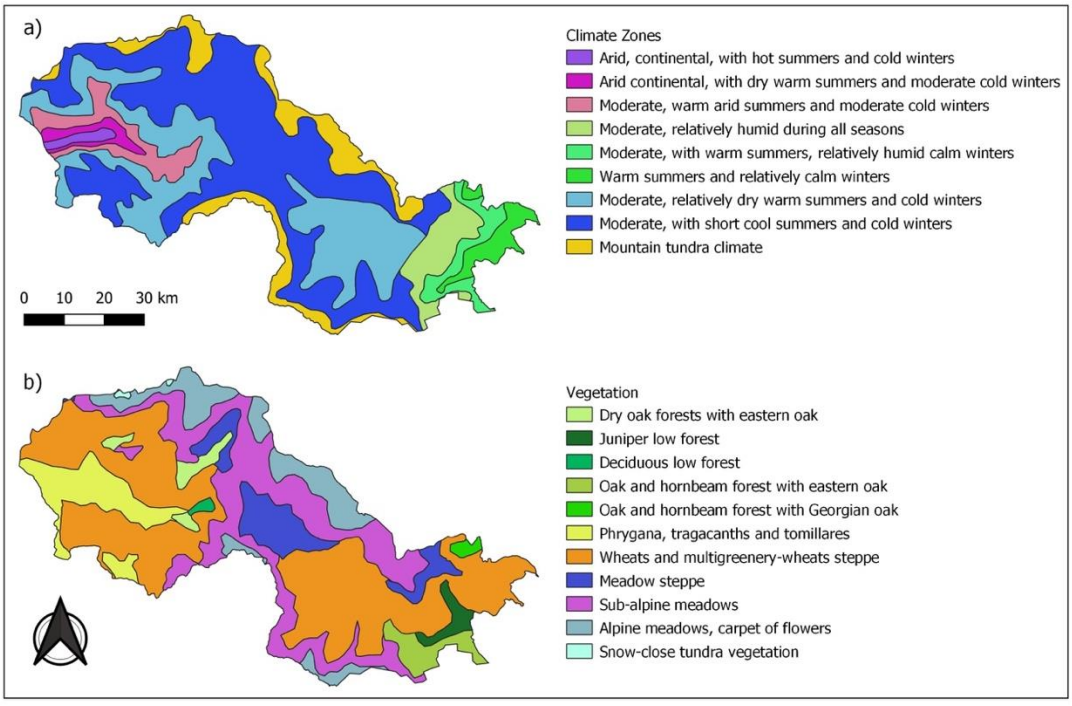


Figure 3 Thematic map of the study region illustrating the climatic and vegetative zonation caused by the orographic setting. Whereas in Vayots Dzor the climate is arid featuring semi-desert vegetation and dry forests the area around Goris is receiving considerably higher precipitation is characterized by deciduous forest and steppe vegetation. With increasing altitude, the steppe grassland is changing to meadow-stepped and (sub-) alpine meadows. (Data: AUA Shapefiles; source: own illustration).

environment. The semi-desert vegetation is characterised by tomillares (i.e. oil-bearing plants of the family Lamiaceae) and tragacanths communities, in which cushion-shaped shrubs dominate. Typical species include wormwood (*Artemisia fragrans*) (figure 4, c) and thyme (*Thymus caucasius*) (Fayvush, Aleksanyan and Bussmann, 2016; Fayvush and Aleksanyan, 2020).

The middle and upper mountain belts (1,200–2,200 m) are characterized by different types of steppe and forest vegetation, meadow-steppes, shrub steppes and tragacanth vegetation. The forest cover in the area is low and restricted to the lower and middle mountain belts. The timberline is at 2,300–2,400 m. The forests in the study region include arid open woodlands in Vayots Dzor as well as oak, hornbeam forests and open coniferous juniper forests in Syunik. Arid open woodlands consist of both coniferous (juniper woodlands) and deciduous species and exhibit a high proportion of endemic species (figure 4, a, b, e & f).

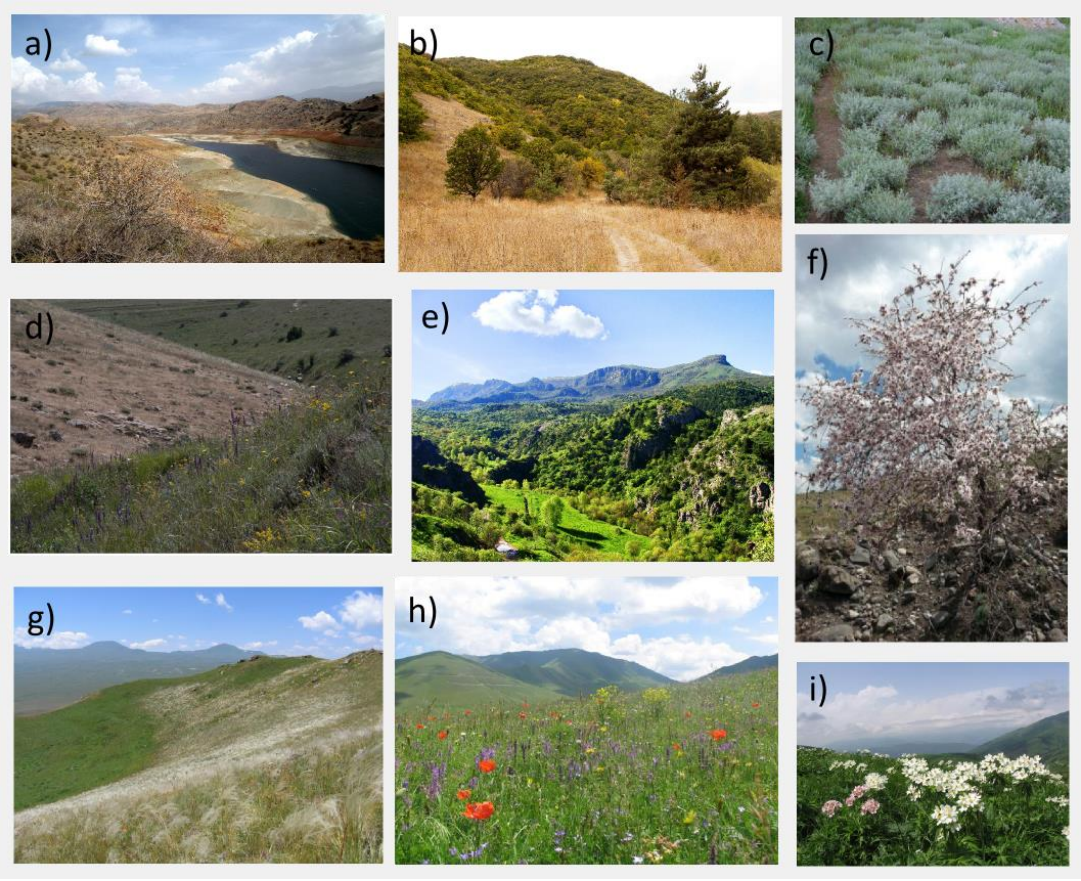


Figure 4 Mountainous landscapes in the study region with specific characteristic plant species composition. a: arid woodland landscape around the reservoir of Herher (*Herher reservoir: low water season*, 2012) *b: sedgy arid woodland close to Jermuk* (Gorbunov, no date)*; c: wormwood (*Artemisia fragrans*)* (Batsatsashvili *et al.*, 2016)*; d: differences in the grassland community for south- and north-facing slopes* (Aleksanyan *et al.*, 2020)*; e: Artavan forest with mixed species* (*Artavan forest*, no date)*; f: wild almond (Amygdalus fenzliana)* (Gulsoy, 2019)*; g-i: feathergrass steppe landscape (g); subalpine meadows (h); Anemone fasciculata (i)* (Asatryan, 2017)*).*

Mountain steppe grassland is the most abundant type of vegetation in the study area occupying all the volcanic plateaus and treeless mountain slopes of the middle mountain belt. One of the most common forms is the feather-grass steppe (figure 4, g). Other forms include koleria grass and wheat grass. The specific community composition depends on the topography and exposition of the slopes as well as the given soil conditions (figure 4, d). Steppes are the most diverse type of vegetation and feature the greatest proportion of endemic species in Armenia (Fayvush and Aleksanyan, 2020). In a recent study the biodiversity of Armenian grasslands was assessed with a standardized sampling and found to be outstandingly high with a vascular plant richness only minimally below the highest record in Transylvania (Aleksanyan *et al.*, 2020). In the upper montane belt, there is a relatively narrow belt of meadow-steppes, in which both steppe and meadow vegetation are associated (Fayvush and Aleksanyan, 2020).

The upland zone (≥2,200 m) is occupied by subalpine and alpine vegetation including subalpine meadows and tall grass vegetation as well as alpine meadows and carpets. The subalpine zone features different types of vegetation such as grass, forbs, mosses or lichens, heathland and scrubby or woody plants (figure 4, h & i). Alpine vegetation is mainly composed by grass-dominated meadows and herb-dominated carpets (Fayvush and Aleksanyan, 2020). Within the study region, large areas are characterized by rocky, bare and sparsely vegetated surfaces with scree accumulations and rock outcrops which are habitat for stenochorous plant species (Fayvush and Aleksanyan, 2020).

The soils within the study region follow the vegetation types and the climatic zones. In the semi-desert Calcisols, Solonetz and Solonchaks are widespread. In the more arid steppes Kastanozems are favoured whereas in more humid steppes Chernozems dominate. Both are rich in organic matter and further benefit from the volcanic substrate. Cambisols are typically found in forested areas. The upland with subalpine and alpine meadows is dominated by little developed soils such as Leptosols and the transitional zone between the subalpine and the steppe zone by Phaeozems (Ghazaryan, 2013).

The location of the study area between the Euro-Siberian and Irano-Turanian phytogeographical regions and between two hotspots of biodiversity (Caucasian and Irano-Anatolian) in particular contributes to the rich biodiversity with a high level of endemism. Wildlife

plays an important role within the mountainous ecosystem and is conserved through designated corridors and protected areas. A current project of the World Wide Fund for Nature and the Ministry of Nature Protection Armenia is the establishment of an ecological corridor to connect the two major protected areas Arevik and Khosrov from which parts are covered by the study region (WWF, no date). This is of significance for one of the most prominent large mammals raising high interest for conservation, the Caucasian leopard (*Panthera pardus saxicolor*). Within recent years the population seems to stabilize with recorded reproduction, evidence of dispersion and active movement of the individuals. One male individual that was born in Nakhchyvan was recorded some 34-36 month later in the Khosrov State Reserve in Armenia, 170 km north-west. The area of this study is situated between those two sites and in direct proximity to vulnerable but viable neighbouring populations (Askerov *et al.*, 2019; Zazanashvili, Manvelyan and Heidelberg, 2020).

Agriculture has a great impact on habitats and biodiversity within the study region. The use of the grassland ecosystems for livestock farming for example is an important issue and requires careful planning of hay cutting and grazing management. Irregular management and lacking organization of hay cutting and pasture grazing is responsible for widespread degradation caused by both over- and under-grazing or -hay-cutting. Disproportional distribution of the pasture load is often the case when distant pastures and meadows suffer from under-use and closer locations from over-use. This results in a change of ecosystems, in particular in the replacement of alpine carpets with alpine, the spread of subalpine weeds into alpine ecosystems and a shift of the lower mountain zone and semi-desert vegetation in the mountain steppe ecosystem (Fayvush and Aleksanyan, 2020). Grassland communities adjacent to pasture areas are also often degraded due to changes in plant composition, erosion and resulting land-slides and mudflows (Fayvush, Aleksanyan and Bussmann, 2016). Animal dung is important for preventing the degradation of pastures and delivers valuable nutrients. However, it is also a free and well-accepted alternative fuel source to firewood and the collection and use is impacting the species composition (GIZ, no date). Finally, intensification and transformation of grassland and other semi-natural sedge-scrub communities into agricultural land can result in both changes in the composition of vegetation and animal communities as well as erosion (Aghababyan *et al.*, 2016).

Threats for the biodiversity associated with crop production include losses of water due to ineffective irrigation and the linked salinization of soils, erosion and pollution by agricultural wastes, lack of crop rotation or violation of ploughing rules. Abandoned fields are widespread in Armenia. They can become a reservoir for weeds or invasive species and form new monodominant communities (e.g. *Silybum marianum*), dispersing and invading the adjacent areas with their seeds and propagates (Fayvush and Aleksanyan, 2020). The on-farm diversity can further be diminished and the distribution of valuable cultural landscapes may decrease. There is evidence that although the biodiversity can increase on abandoned farm fields in an early succession stage, it may decline in later succession stages. Shrub encroachment can serve as fuel for wildfires. Abandonment may further cause spill-over effects that lead to the marginalization of historic agricultural landscapes (Elbakidze and Angelstam, 2011; Prishchepov *et al.*, 2012; Yin *et al.*, 2018). This emphasizes the importance for an effective use of farmland in Armenia (CBD, 2014).

As illustrated in table 1, agriculture in the study area is primarily smallholder-based with family farms and private households engaging in different forms of cultivation or animal husbandry activities. Although mixed farming activities are very common, livestock farming and pastoralism is of particular importance for the smallholders in the study area. Especially in the area of Sisian-Goris most of the farmers are raising livestock. Together with the district of Gegharkunik, Syunik has the largest number of sheep from all administrative regions in Armenia. Many of the small-holders are transhumant pastoralists. In the late spring they move their livestock to summer pastures with small stations where they live several weeks or months. With the first snow in the autumn they bring them to their winter pastures in the valleys where the herders live. In order to feed the cattle in the winter, hay meadows are cut in July and August and hay or silage is prepared (see figure 5).

Table 1 Characteristics of agriculture in Armenia. Number of farms and family farms with the proportions of family farms employed in specific farming systems for the administrative districts of Vayots Dzor and Syunik (FAO, 2020b).

	Total number of farms	*Family Farms*	*Land cultivation only*	*Livestock farming only*	*Mixed activities*
Vayots Dzor	10 384	8 932	2 114	379	6 439
Syunik	21 361	19 989	9 369	727	9 893

Figure 5 Beehives and hay transport in the mountainous plateau of Syunik. Sources: left: (Feminó, no date)*; right:* (Davtyan, no date)*.*

Besides the significance of livestock in Syunik, a considerable proportion of smallholders are cultivating their land with crops only. Although the smallest size class is the most abundant in this region, the farm size distribution also shows a share of more than 10% of farms with 5 ha or larger (see figure 6). Apart from Shirak, this is the highest proportion of the biggest farm size class in all the Armenian regions. The smallholders in Syunik primarily cultivate wheat and barley, leguminous crops, potatoes and vegetables as well as forage crops for their livestock. Beekeeping is an important farm activity in Sisian-Goris (figure 5). With about 31,000 bee hives Syunik is the region where most smallholders practice beekeeping (FAO, 2020b).

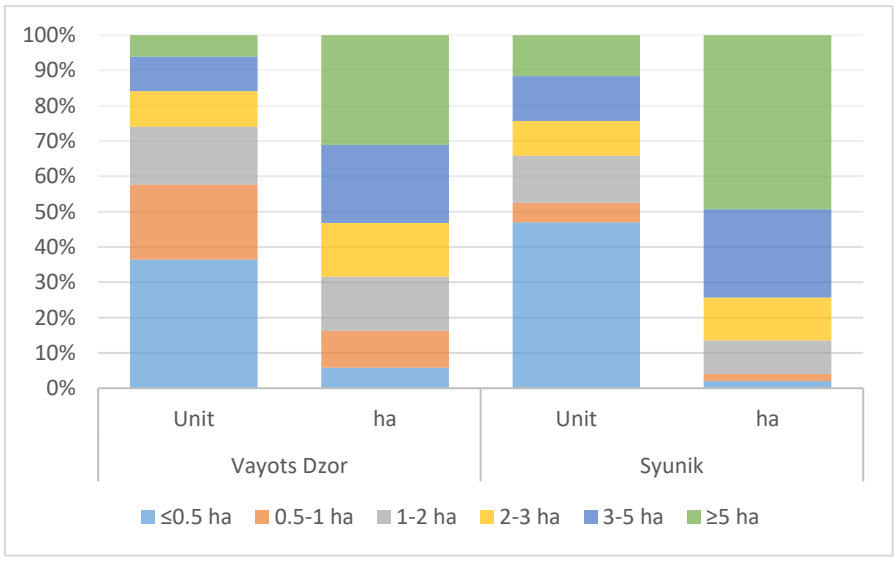


Figure 6 Farm size distribution for Vayots Dzor and Syunik with farm unit size classes and associated plot size proportions. Compared to Vayots Dzor, Syunik features a considerably higher proportion of larger fields (FAO, 2020b)*.*

Perennial crops and orchards in particular play a significant role for the livelihoods of the smallholders in both Vayots Dzor and Syunik with 1,479 ha and 1,326 ha area of plantings respectively. Important trees include apricot and other stone fruits, pome fruits and almonds. Grape vines and vine yards are an important source of income for many smallholders in Vayots Dzor with a total of 805 ha of grape vine plantings. It should be mentioned that this number only accounts for smallholder farms and not for larger commercial grape vine farms in Vayots Dzor. Besides perennial plantings the smallholders in Vayots Dzor cultivate mainly grain crops, forage crops, vegetables and melons but also industrial crops such as tobacco and sunflower. Livestock plays also an important role but is mostly included in mixed farming activities as opposed to Syunik (FAO, 2020b).

The diversity of wild plants and crop relatives is of significantly value for the rural population in the study region. Old individuals of threatened wild pear trees are for example valued as a symbol of the village of Herher in Vayots Dzor (figure 7) (Asatryan, 2019). The wild plants are further intensively used for food, fuel and construction material as well as medicinal and ornamental purposes. There are about 200 edible wild plants in Armenia which are used in both fresh and processed forms comprising up to 10-15% of the total diet (Fayvush, Aleksanyan and Bussmann, 2016). In a recent study Nanagulyan *et al.* (2020) assessed wild plants and fungi sold in the farmer markets of Yerevan. Over four years they found 163 plant species of which some had never been recorded before.

Figure 7 The oldest individual of the endangered wild pear Pyrus gergerana located besides the road to Herher (Asatryan, 2019).

2.2.LULC class typology

This study focuses on agricultural land-use systems but also includes other land cover types in the area. The typology aims to deliver a coherent set of LULC classes with both agricultural target classes (1-6) and rich semi-natural by-product classes (7-11). The use of a synoptic and obscure "other" class is avoided. A LULC classification requires the definition of class boundaries that should be clear, precise and based on objective and possibly quantitative criteria. The LULC typology used in this study is therefore described in the following (the class abbreviations used in figures and tables are given in brackets):

1) **Grassland/ potential pastures (GL):** Non-woody herbaceous vegetation and herbaceous vegetation grazed by livestock; shrubby woody vegetation canopy cover of less than 30-40%. The phenology is characterized by the vegetation period of the given altitudinal zone and herbaceous species.
2) **Hay meadows (HM):** Non-woody herbaceous vegetation which is harvested once during the vegetation period for hay making. The meadows are rainfed and agricultural inputs and management such as irrigation, fertilization as well as tillage are absent.
3) **Improved meadows (IM):** Intensively managed herbaceous vegetation that is irrigated and harvested at least twice during the vegetation period. The irrigated meadows are mostly found within riverine floodplains in the valleys.
4) **Field crops (FC):** Land covered by annual crops followed by a harvest and a fallow period. Examples for crops are: cereals, oil seeds, vegetables and root crops and leguminous forage crops. This class does not include grass and meadow forages. Some of the crops are irrigated (e.g. vegetables) while other crops are mainly rainfed (e.g. grains and forages). It also comprises tillage or ploughing and possibly agricultural inputs like fertilization.
5) **Vine yards (VY):** Vine yards with a typical row planting structure. This class comprises both larger commercial vine production areas as well as smaller subsistence-based yards within settlements. They may be irrigated and the herbaceous layer below the vines may be mowed. The permanent vegetation is characterized by a full phenological cycle from the sprouting of leaves in spring to the fruit ripening in summer to the leaf senescence in autumn.

6) **Fruit orchards/ home gardens (FO):** Fruit and nut trees within orchards or home gardens. The canopy cover in the case of orchards is dispersed and ranges from around 10-70% depending on the age of the trees. Besides the larger orchards it also comprises fruit and nut trees within smaller home gardens in settlements. Fruit orchards may be irrigated and the herbaceous layer below the trees may be mowed. The permanent vegetation is characterized by a full phenological cycle from the sprouting of leaves in spring to the fruit ripening in summer to the leaf senescence in autumn.
7) **Shrubland (SL):** semi-natural grassland vegetation dominated by shrubby, woody species. The shrubby canopy cover ranges around 60-70%.
8) **Forest/ woodland (WL)**: Trees higher than 2 meter and with a canopy cover of more than 80%. This excludes trees outside of forest such as fruit trees and nut trees in orchards and home gardens but includes urban trees with the defined minimum tree height and canopy cover. Depending on the species, the phenology can be characterized by evergreen or deciduous traits.
9) **Bare ground/ sparse vegetation (SP):** Surfaces with a vegetation cover of less than 20% or below a vegetation index threshold given the climatic and ecological characteristics of the area. It includes bare rock and boulders in the mountainous landscape, river sediments, open soil surfaces due to disturbances or hazards such as landslides and bare rock surfaces of mining areas. It also accounts for overgrazed bare soils on grasslands and desert soils with a very low vegetation cover. It does not include fallows on agricultural land.
10) **Surface water (W):** Water bodies larger than 10 m x 10 m and open to the sky with no canopy cover.
11) **Urban/ artificial (U):** Urban structures such as paved roads, buildings and other urban features. Industrial structures include factories as well as power plants and solar plants. Concrete structures of hydropower plants and mining operations are also included.

2.3.Overview of the LULC classification scheme

This section provides an overview of the general classification scheme of LULC as applied in this study (see figure 8). For the supervised classification, a pixel-based approach was chosen due to its simplicity and practicality for classification and accuracy assessment. It is widely applied and often yields similar accuracies as object-based approaches (Olofsson *et al.*, 2014; Fu *et al.*, 2017; Berhane *et al.*, 2018). The classification follows the classic approach of classifying all classes at a time as opposed to hierarchical or sequential approaches.

The supervised classification is based on a stratification into two agroecological zones or sub-regions: 1) the district of Vayots Dzor (Yeghegnadzor and Vayk), and 2) the administrative sub-districts of Sisian and Goris (see figure 2). The rationale for this were the pronounced climatic differences resulting in spectral variability within classes on the one hand, as well as the limitations of training sample sizes as specified by the GEE quota on the other hand. Initial classifications for the whole study areas delivered unsatisfactory map predictions with apparent

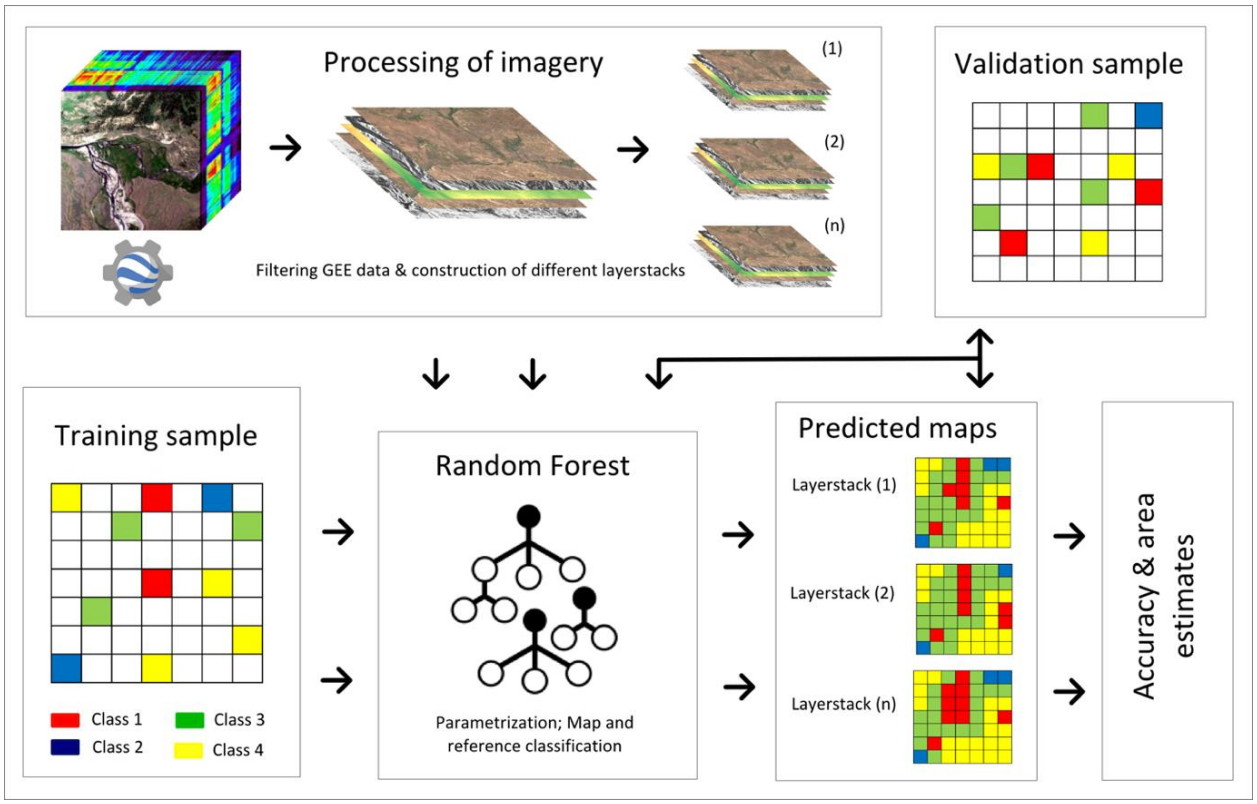


Figure 8 Schematic overview of the classification system. The RF classification follows a pixel-based approach in a classic non-hierarchical fashion. The imagery and ancillary data is processes within GEE and different datasets (layerstacks) are produced. Based on the dataset and the training sample RF predicts different maps and the mapping accuracy is determined using an validation sample that is independent from the training sample (source: own illustration).

class confusion. In order to account for this, the training data was extended to the point where the stratification was necessary because the training sample size exceeded the GEE quota for classification. With a stratification into two sub-regions these limitations could be overcome. This also implies that the RF map prediction and the accuracy assessment are produced separately for the two sub-regions.

The remote sensing imagery was classified with the RF machine learning algorithm implemented in GEE. Training and validation data were sampled as input for the RF classification and the accuracy assessment. The remote sensing data was pre-processed within GEE and different datasets (image layerstacks) were constructed (see section 2.5). Both the datasets as well as the training and validation samples were then used by RF for the different map predictions. The output comprises a map and a reference classification for each dataset that is used to assess the accuracy by constructing an error matrix from both outputs. Sample based accuracy and area estimates were derived for underlying statistical support of the map predictions. In the following sections these steps are described briefly.

2.4.Satellite data and sensor description

This section describes the remote sensing imagery, the sensors as well as the ancillary data used for both as input for the LULC classification and as reference data for the acquisition of validation samples. The data products are further linked to the specific use in this study and the platforms used to display and obtain the data are associated. The remote sensing data includes: (1) Sentinel-2 MSI Level-2A images; (2) Sentinel-1 SAR imagery; (3) Shuttle Radar Topography Mission (SRTM) digital elevation image; (4) Visible Infrared Imaging Radiometer Suite (VIIRS) nighttime lights image, and (5) PlanetScope and RapidEye imagery.

The Sentinel-2 earth monitoring constellation includes two polar-orbiting satellites launched in mid 2015 and early 2017 as part of the Copernicus Land Monitoring program. The satellites are in the same sun-synchronous orbit and phased at 180° to each other. They carry a wide-swath (290 km), high-resolution, multi-spectral imaging instrument (MSI) which captures 13 spectral bands with a spatial resolution ranging between 10 and 60 meters. Together the sensors provide nearly global coverage with a revisit frequency of approximately five days between latitudes 56° south and 84° north (ESA, 2015).

The Sentinel-2 imagery was accessed in the GEE cloud environment as a readily pre-processed product. The elementary level of Sentinel-2 MSI products are granules of a fixed size dependent on product level. The image product available in GEE is divided into 100 km tiles in UTM/WGS84 projection. The Level-2A data was used in this study because it offers the highest degree of image pre-processing. It includes a scene classification and an atmospheric correction applied to the Top-Of-Atmosphere Level-1C orthoimage product. The main output is an orthorectified Bottom-Of-Atmosphere corrected surface reflectance product (ESA, 2015). The applied algorithm corrects for both atmospheric and topographic effects (ESA, no date a).

The study region is covered by four of the Sentinel-2 footprints: 38TMK, 38TNK, 38SNJ and 38SPJ. All available Level-2A imagery for the period between the 1st of September 2018 and the 30th of April 2020 were considered for the LULC classification and processed further within GEE. This resulted in 808 images for the given footprints. Figure 9 illustrates the cloud cover proportions of the included Sentinel-2 images.

The Sentinel-1 constellation consists of two polar-orbiting satellites which operate day and night. The satellites are capturing C-band images with a SAR instrument. The active sensors are able to acquire images in any light or weather conditions because they are operating at wavelengths not interfered by cloud cover or a lack of illumination. The instrument is operating

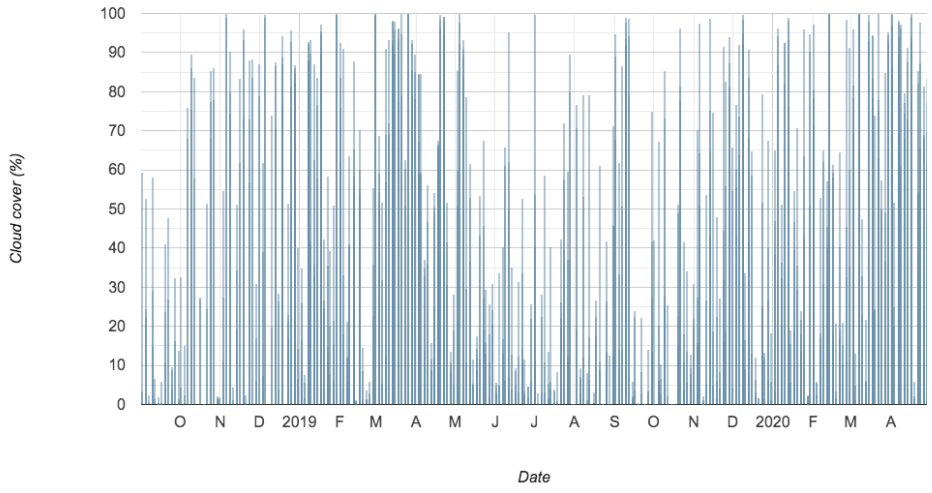


Figure 9 Bar chart showing the cloud cover in percentage of the image area for all Sentinel-2 observations considered in this study. Within the winter and spring months the cloud cover is rather high whereas in the summer there is higher availability of clear observations. The graph was produced in GEE using the stored cloud percentage metadata.

in four exclusive imaging modes with different resolution (down to 5m) and coverage (up to 400 km) (ESA, no date d). The C-SAR instrument allows operation in both single polarisation (VV or HH) and dual polarisation (HH+HV, VV+VH). For each observation detailed metadata of the satellite position and attitude is available. The main conflict-free mode over land is the Interferometric Wide Swath Mode (IW), with VV+VH polarisation (ESA, no date c).

Sentinel-1 data was also obtained as a pre-processed product within the GEE platform (GEE, no date c). For this study the Level-1 Ground Range Detected (GRD) product was used. It includes focused SAR data that has been detected, multi-looked and projected to ground range using an earth ellipsoid model. In contrast to the Single Look Complex (SLC) product, the GRD product has almost square pixels and pixel spacing with reduced speckle at the cost of spatial resolution and the phase information is lost. The GRD product is available in three different spatial resolutions: (1) Full Resolution (FR), (2) High Resolution, and (3) Medium Resolution (MR) (ESA, no date b). In the case of IW mode, the GRD product is available for high and medium resolution. The scenes of the product have an additional 'angle' band that stores the approximate viewing incidence angle in degrees at every point (GEE, no date c).

The pre-processing steps for the calibrated and ortho-corrected GRD product include a thermal noise removal, radiometric calibration and a terrain correction using digital elevation products. The calibrated and corrected values are finally converted to a backscatter coefficient (σ°) in decibels (dB) using log scaling (10*log10(σ°)) indicating whether the radiated terrain scatters the incident radiation preferentially away from the sensor (dB < 0) or towards the SAR sensor (dB > 0) (GEE, no date b). Each scene of the collection in GEE is accessible in the different spatial resolutions, four band combinations (corresponding to scene polarization) and three instrument modes (IW, EW, WV). The use of the collection for the creation of best pixel composites requires filtering down to a homogeneous set of bands and parameters (see section 2.5). For the purpose of this study, the IW mode with VV+VH polarisation with a high spatial resolution was chosen. Scenes between the 1st of April and the 31th of October 2019 were considered to complement the optical data in the LULC classification. This resulted in 87 SAR scenes covering of the study region.

The ancillary data used in the study includes a digital elevation model of SRTM and the VIIRS nighttime lights product, both partially acquired and made available by NASA. The two products

are also directly available within GEE in a pre-processed form. The new digital elevation product of SRTM was released in 2015 with a spatial resolution of 30 m. It provides near global land surface coverage between +/- 60° of latitude (Farr *et al.*, 2007). The SRTM product in GEE was pre-processed by filling voids using open-source data (SRTM, 2015). VIIRS is one of five instruments onboard the Suomi National Polar-orbiting Partnership (NPP) mission launched in 2011. VIIRS acquires a panchromatic Day/Night band (DNB) that is ultra-sensitive under low-light conditions and therefore supports the observation of nighttime lights with a high spatial and temporal resolution. The nighttime lights imagery is also directly accessible within GEE as monthly averaged radiance composites of the DNB. In this study Version-1 of the product was used. The prrprocessing includes a cloud cover reduction using the VIIRS Cloud Mask product and exclusion of data close to the swath edges. It is filtered for stray light, lightnings, lunar illumination and cloud cover (NASA, no date). For this study the monthly composites for the entire year of 2019 were processed within the GEE environment and included in the input dataset for classification.

VHSR images from Planet Labs Inc were used for the assessment of reference data. Every day the company acquires images from more than 160 dove satellites and downlinks 340 million square kilometres of data with near global and daily coverage (Collison, 2017). Planet operates the PlanetScope and RapidEye Earth-imaging constellations with a constantly improving on-orbit capacity. Both imaging constellations acquire images with four bands between 455 nm and 860 nm wavelength (RGB and NIR). The spatial resolution is 3 m for the PlanetScope and 6.5 m for the RapidEye constellation (Planet, 2016). Via the Planet Explorer platform quarterly and monthly surface composites as well as unprocessed daily products could be accessed. The surface reflectance composites are corrected for atmospheric effects and calibrated for sensor characteristics and sensor-to-sensor variability, resulting in consistent imagery data across seasons, continents and instruments (Collison, 2017).

2.5.Processing of remote sensing imagery

The pre-processed remote sensing data described above was further processed within the GEE cloud environment to construct different layerstacks used as input for the RF classifier. Major pre-processing steps included: (1) filtering the data for start and end-dates, the region of interest (ROI) and other parameters (e.g. bands of optic sensor / swath mode & polarization for SAR data);

(2) cloud masking of the filtered optical image collection; (3) Calculation of spectral derivates for each filtered optical image and inclusion as image band in the image collection (e.g. normalized difference vegetation indices); (4) Calculation of spatial-temporal metrics from the entire filtered optical image collection (e.g. min, max, mean, standard deviation of each band pixel in all filtered images); (5) Construction of cloud free monthly median composites of the optical image collection to construct a monthly time series between April and October; (6) Construction of monthly composites from means at different polarizations and look angles of the filtered SAR data to construct a monthly time series between April and October; (7) Calculation of texture metrics for a summer median composite (July-September) of both the filtered optical image collection (NIR) and the filtered SAR data; (8) Calculation of topography metrics from the SRTM digital elevation data (i.e. slope, aspect, eastness, northness); (9) calculating a yearly median composite from the monthly nighttime lights composites. Each of these steps and the underlying rationale is shortly described in the following. The final construction of the different input layerstacks for RF is described and illustrated subsequently.

First of all, the data was filtered for the start and end dates and certain sensor parameters or acquisition bands. In the case of the optical data the temporally filtered collection contained all images within the1st of September 2018 and the 30th of April 2020 (808 images). The spectral bands of each image were filtered for: blue (band 2), green (band 3), red (band 4), NIR (band 8A), short wave infrared (SWIR)-1 (band 11) and SWIR2 (band 12). The SAR data was filtered for all available images acquired in IW mode with VV+VH polarisation between the 1st of April and the 31st of October 2019 (87 images). Finally, the monthly nighttime lights composites were filtered for 2019 and average radiance. All images were clipped with the shape of the study region.

The cloud cover of the optical image collection was masked using the quality band of the product. Both opaque clouds and cirrus clouds were masked out with a buffer of 90 pixels around the cloud cover. The filtered and cloud masked optical image collection was then used to calculate several derivates of the spectral bands of the image collection. For the classification of agricultural land use and land cover by-products different indices and spectral transformations for vegetation, bare surfaces and urban structures were considered. The multiple indices for vegetation vary in terms of adjustments to soil and moisture conditions, canopy structure or background signals.

The derivates were calculated and included as bands in the optical image collection. The table of appendix I shows a list of the derivates with the formulas and a short description.

The entire optical image collection with the added derivates was then used as temporal bin to derive spectral-temporal phenology metrics. Phenology metrics are simple descriptive statistical metrics which describe a distribution of values - in this case the temporal distribution of pixel values for the different bands. The descriptive statistics include: minimum, maximum, mean, median, standard deviation, 25th quartile and 75th quartile. The mean band value of a certain pixel is calculated from all the pixels within the defined temporal range (1st of Sep 2018-30th of April 2020). The metrics basically describe change over time and for spectral bands or derivates that are sensitive to vegetation, changes during the phenological year. Similarly, other land surfaces may also change more or less during the year. The temporal changes of bare surfaces and soil signals for example are important for agricultural land use. Using an extended temporal bin of more than one year, aids in capturing land-use practices such as tillage and ploughing, typically taking place in autumn and spring. For this temporal bin, the metrics were calculated using all optical bands and the derivates and resulting in 119 band metrics which were then assembled in the main layerstack.

The cloud coverless monthly median composites were calculated following best (available) pixel compositing. Remote sensing data does not rely on a single image but can be composited by multiple images following certain quantitative criteria. Best pixel composites include quality composites such as greenest or barest pixel composites but also mean or median composites for a specific temporal bin in order to construct a time series for example. The process is therefore based on the same concept as the phenology metrics. Best pixel compositing was used to construct a time series displaying both agricultural management and the phenological changes of the vegetation in the study area. For this a temporal bin of one month was considered as sufficient. Although the temporal resolution allows composites with shorter time frames, there is also substantial cloud cover for certain periods which limits the data availability (see figure 9). Thus, monthly median composites were calculated using optical data as well as derivates related to vegetation or bare soils for seven consecutive months (April-October). The resulting monthly time series was then included in the main layerstack.

Because of the heterogeneity of the filtered SAR image collection regarding the image acquisition parameters, it was necessary to filter the data for ascending and descending modes and calculate the composites from means at different polarizations and look angles (GEE, no date b). This resulted in two mean composite bands for VV and VH with combinations of polarizations and angles for each month (April-October). The monthly SAR composites were then also included in the layerstack.

Due to its practicality, the Grey level co-occurrence matrix (GLCM) was chosen to be included as texture for the LULC classification. As output, a continuous measure of spatial information is provided which can be used for further processing. The metrics can be separated in interior and edge textures. The choice of the set of measures depends on the application and specific objectives (Hall-Beyer, 2017). Based on Hall-Beyer's (2017) analysis of GLCM texture performance for different land cover and surface types four texture metrics were considered to be included: two interior texture metrics (correlation and variance) as well as two edge texture metrics (entropy and contrast). Two sets of these texture metrics were derived for both the optical and the SAR imagery for a three-month median composite in summer (July-September). In the case of the optical data, NIR was used to derive the GLCM textures. For the SAR imagery the two mean angle and polarization composites were used to derive texture metrics. This resulted in four optical texture and eight SAR texture bands that were added to the layerstack.

Other ancillary data includes the SRTM digital elevation product and the VIIRS nighttime lights product. The digital elevation image was used to derive four topographic continuous metrics, namely the slope, the aspect, eastness and northness. Slope and aspect are the degree of change of elevation in magnitude and orientation for the steepest descent vector, respectively. Eastness and northness are the sine of the slope, multiplied by the cosine and sine of the aspect, respectively (Amatulli *et al.*, 2018). The original elevation band as well as the derived topography metric bands were included in the layerstack (five bands). In case of the nighttime lights a median composite was calculated using the 'avg_rad' band for the entire year of 2019 (12 month). The resulting image was included in the layerstack.

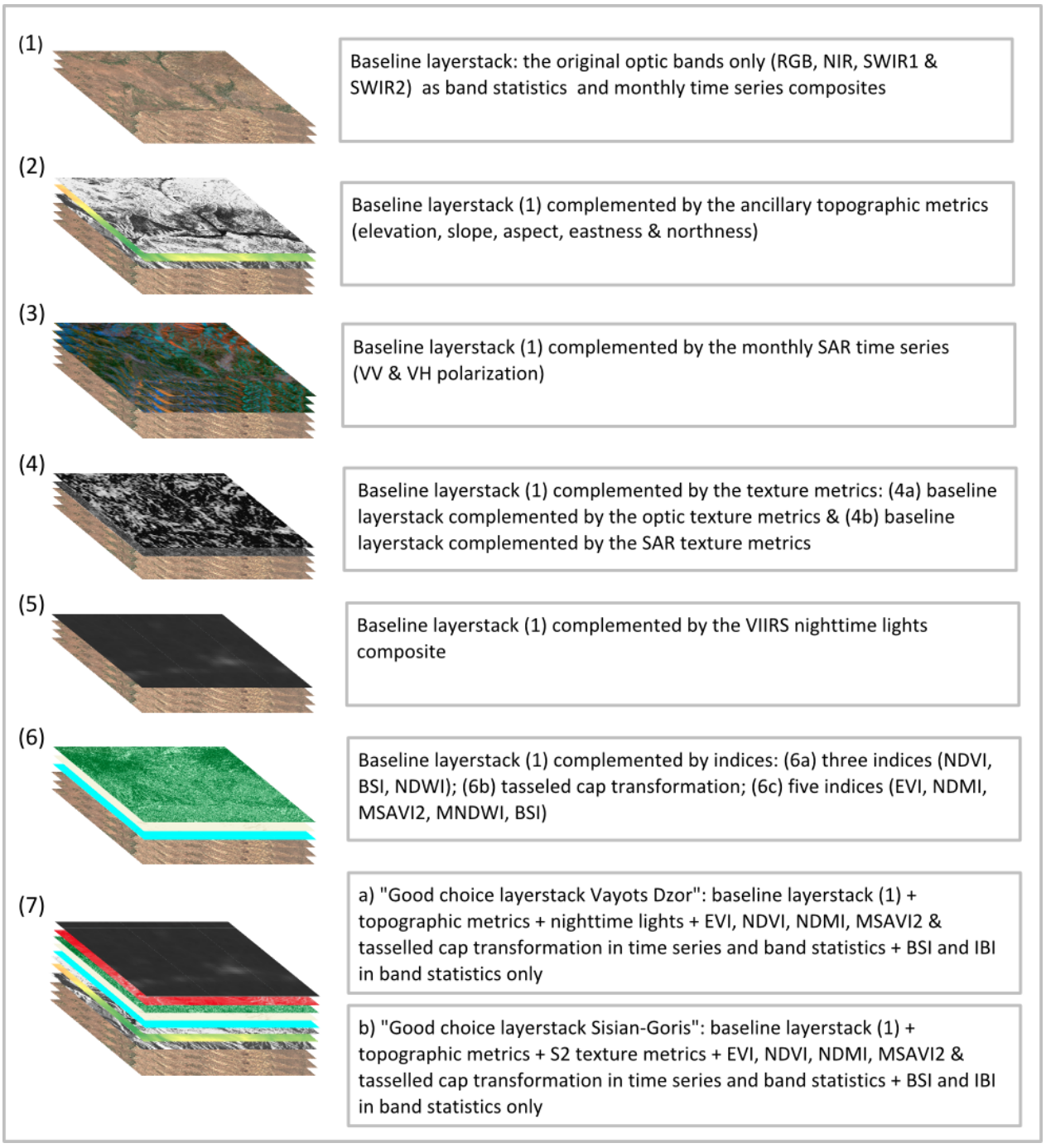


Figure 10 Overview of the different layerstacks for the RF classification (source: own illustration).

The digital elevation image was used to derive four continuous topography metrics, namely the slope, aspect, eastness and northness. Slope and aspect are the degree of change of elevation in magnitude and orientation for the steepest descent vector, respectively. Eastness and northness are derived from the sine of the slope, multiplied by the cosine and sine of the aspect, respectively (Amatulli *et al.*, 2018). The original elevation band as well as the derived topography

metric bands were included in the layerstack (five bands). A median composite of the VIIRS nighttime lights was further calculated using the 'avg_rad' band for the entire year of 2019 (12 months). The resulting image was included in the layerstack.

The main layerstack containing all the processed bands was used to construct different datasets for the random forest classification (see figure 10). First, a baseline dataset was constructed containing spatial-temporal metrics and a monthly time series of the original optical bands. In order to evaluate how the LULC classification changes with different additional data this baseline dataset was then complemented with: (1) topography metrics; (2) the monthly SAR time series; (3) the texture metrics derived of the optical as well as the SAR composite; (4) different sets of indices and (5) the nighttime lights composite. Based on the observations of the different RF models, a “good choice layerstack” with data that improved the classification results was constructed and customized for each of the sub-regions (layerstack 8). The added value of the multisource data and considerations for the “good choice” dataset can be obtained from the results.

2.6.Random forest

Random forest is an ensemble classifier that produces multiple decision trees using a randomly selected subset of training samples and input variables. The different decision trees are grown by drawing a subset of training samples through replacement (bagging) meaning that the same sample can be selected several times whereas other samples may not be selected at all. The bag fraction defines the proportion of samples used to train and construct the trees (in-bag sample) and samples left out for an internal cross-validation to estimate how well the resulting RF model performs (out-of-bag error) (Belgiu and Drăgu, 2016). For the classification of LULC the GEE default bag fraction of 0.5 was used.

The decision trees are independently produced without pruning and each node is split by the number of features (*Mtry*), selected at random (bootstrapped sample). The number of features is usually the square root of the number of input variables. The forest of random trees is grown up to a user-defined number of trees (*nTree*) to create trees with a high variance and low bias. There is evidence that the empiric value of 500 trees is suitable for most remote sensing applications and was therefore used within this study. The final classification decision is

determined by averaging (arithmetic mean) the class assignment probabilities calculated by all produced trees. The result is an unlabelled data input that is evaluated against all decision trees created in the forest and each tree votes for a class membership. The membership class with the maximum votes is the final class selected by the RF algorithm (Belgiu and Drăgu, 2016).

RF addresses most of the challenges related with supervised classifiers. It can deal with a high dimensionality of input data and is able to classify multi-source data. Further, it shows a low sensitivity to the quality of training data and overfitting. However, with imbalanced training samples RF tends to favour the most representative classes (Belgiu and Drăgu, 2016). GEE supports the calculations of RF variable importance. The variable importance is relevant in many applications in order to select those variables with the greatest ability to discriminate between the classes and reduce dimensionality. However, a reduction of variables was not applied because it was not clear how the implemented algorithm calculates the importance of the variables.

2.7.Acquisition of training and validation data

Supervised image classification requires the sampling and class allocation of training and validation data to predict and map thematic features of interest and to assess the accuracy of these predictions (Foody, 2004).

Data sampling

According to Olofsson et al. (2014), sampling strategies can be grouped into non-random and random designs. They are defined in terms of inclusion probability, which relates to the likelihood of a given unit to be included in the sample (Stehman, 2000). A random sampling needs to meet two conditions: (1) the inclusion probability must be known for each unit selected in the sample and (2) the inclusion probability must be greater than zero for all units in the region of interest (Stehman, 2001). Ideally, both training and validation data should be obtained by random sampling. However, a random sampling approach can be time consuming and difficulties may arise in the case of a required ground truth assessment, given e.g. accessibility of the locations in the study region. Further, randomly sampled training data may not adequately reflect the spectral variances of the mapped classes in heterogenous landscapes. In these cases, a non-random sampling approach may be chosen, characterized by purposely and opportunistically allocating units to the sample. Nevertheless, if training and validation data are non-randomly sampled, the

data violates the assumption of independence and optimistic bias with overestimations of predicted accuracies is likely to affect the results (Hammond and Verbyla, 1996; Zhen *et al.*, 2013).

Training and validation data should reflect the proportions of class representation on the ground. Unequally or imbalanced proportions of training or validation sites relative to the actual ground proportions impact map predictions and accuracies. Classifiers are then likely to favor the 'majority' classes within the training data and accuracy results may be biased. For training data, over-represented classes may dominate the predicted map while under-represented classes may be marginalized. In order to work with imbalanced datasets or cases where a class exhibits a very small proportion (i.e. rare classes), over- and under-sampling can be applied to construct more balanced training or validation datasets (Millard and Richardson, 2015).

Following Waldo Toblers (1970) first law of Geography, "everything is related to everything else, but near things are more related than distant things", spatial dependence might greatly affect the classification results. Spatial autocorrelation is an important measure of the spatial structure of a variable. A non-random spatial pattern can either show positive or negative spatial autocorrelation (Spiker and Warner, 2007). Spatial autocorrelation should be considered when constructing the datasets for both training and validation. A commonly used approach when allocating training samples is the use of polygons to delineate areas of known class membership in the image. Millard and Richardson (2015, p. 22) note that "[…] this method produces a highly clustered training sample with inherently high spatial autocorrelation." A clustered training sample (i.e. a large proportion of points of a certain class in close proximity) could lead to a strong positive spatial autocorrelation and surrounding pixels of a different class are likely to be misclassified. Correspondingly, clusters of validation sites for a certain class can be responsible for inflated classification accuracies and biased area estimates.

In the case of training and validation sampling, spatial autocorrelation will be high at a smaller distance from the objects (i.e. neighboring pixels) and will increase with some distance. Although in some cases (e.g. for rare classes/ clustered land use in valleys) spatial autocorrelation may not be avoided completely, it should be reduced as much as possible and occurring autocorrelation should be reported. A widely used metric to assess spatial autocorrelation is Moran's I statistic (Eq. 1) (Moran, 1950),

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \qquad \textit{(Eq. 1)}$$

where N is the number of spatial units indexed by *i* and *j*, *x* is the variable of interest, w_{ij} is a set of corresponding weights and $\bar{x}$ is the sample mean. The values of I usually range from -1 to +1 with negative values indicating negative, positive values indicating positive and the value of zero indicating no autocorrelation.

Class allocation

The sampling of training and validation datasets requires the identification of the reality on the ground (i.e. the specific type of land use or land cover found in a place at a certain point in time). This could be achieved by acquisition of ground truth through field visits, detailed thematic spatial data or by interpretation of imagery, in particular very VHSR imagery. Approaches may also be combined in order to reduce limiting factors or to make use of beneficial synergies. While collecting ground truth trough field visits is the ideal, time, costs and accessibility can limit such endeavours.

Based on these considerations concerning the quality of training and validation data for remote sensing classifications, a dataset of both training and validation sites was constructed within this study. In the following, the data sampling and class allocation methods are displayed in depth, illustrating the process in a transparent fashion in order to make it reproducible for other researchers. Limitations and problems encountered during the sampling protocol are also displayed.

Training data collection

In a training stage, areas of known class membership are identified in the image (training sites). The remotely sensed response of these training sites is then characterised from the sample of pixels they contain. The training data is used to train the random forest classifier. The training statistics are used to allocate each pixel in the image to the class with which it has the greatest similarity. The training data is thus a mean to extract thematic information from the remotely sensed imagery and direct the classification (Foody and Mathur, 2004). The accuracy of the map

prediction is crucially dependent on a variety of pre-class allocation issues and in particular the training stage (Foody, 2004; Foody and Mathur, 2004; Millard and Richardson, 2015).

Training data sampling

Unfortunately, no accurate spatial data was available for the study region that captured the agricultural land-use classes of interest. To estimate balanced training sample sizes, official cadaster data provided by the Armenian government, translated to English, served as a first estimation of the actual land-use proportions in the study region. Since the data was structured for the different administrative districts and for different use and ownership categories, the area information was extrapolated and adjusted for the study region and classes of interest. This extrapolation is again subject to inaccuracies and can only serve as a broad estimation of training data proportions. The resulting area proportions of the LULC classes in the two zones of the study region based on the land balance is shown in table 2.

A non-random sampling design was chosen to allocate single-pixel training samples. Although a random sampling design is generally preferable, it was rejected due to the work load, the given circumstances in respect to reference data and in particular to adequately represent the spectral variances of the classes, which is of special importance regarding the spectral similarity of the target classes. Olofsson et al. (2014) highlight the importance to define a minimum mapping unit (MMU) when allocation training and validation samples. The definition of the smallest observable feature that can be captured reliably depends on both the technical possibilities of the source data (i.e. spatial resolution) and the class or feature of mapping interest and can have important implications for the accuracies of thematic maps (Herold, 2011). Although it is largely agreed that the smallest observable feature that can be identified needs to be four contiguous pixels in size (in the case of Sentinel-2, 40 m x 40 m), an MMU of one pixel was chosen (20 m x 20 m) to allow the detection of smaller agricultural land-use patches common in the study area.

Table 2 LULC proportion as derived from the government land balance and proportions of training samples for both zones of the study region.

LULC Class	LULC proportions (land balance)				Training samples (pixel count)			
	Vayots Dzor		*Sisian-Goris*		*Vayots Dzor*		*Sisian-Goris*	
	Area (ha)	% of total	Area (ha)	% of total	Area (pixel)	% of total	Area (pixel)	% of total
Grassland	119005,51	50,83	119375,38	52,58	2055	25,99	2421	25,76
Hay meadows	51781,26	22,12	33198,18	14,62	1208	15,28	1749	18,61
Improved meadows	8630,21	3,69	5533,03	2,44	387	4,89	284	3,02
Field crops	19259,40	8,23	32266,84	14,21	712	9,00	1633	17,37
Vine yards	893,30	0,38	29,94	0,01	587	7,42	n.d.	n.d.
Fruit orchards	1011,10	0,43	423,92	0,19	724	9,16	660	7,02
Shrub land	7544,15	3,22	10012,69	4,41	958	12,11	1082	11,51
Forest/ Woodland	14697,75	6,28	17943,82	7,90	322	4,07	502	5,34
Sparse vegetation/ bare	2346,00	1,00	303,44	0,13	448	5,67	457	4,86
Water	3297,43	1,41	1925,86	0,85	94	1,19	148	1,57
Urban/ Build up	5655,40	2,42	6012,41	2,65	413	5,22	464	4,94
Total	234121,50		227025,52		7908		9400	

For both sub-regions, a total of 17308 pixels have been sampled as training data by allocating points in an opportunistic and non-random fashion (see figure 11), covering about 0.15% of all pixels in the study area. Table 2 shows the allocation of training samples, with pixel numbers per class and the associated proportions in detail for both zones of the study area. During the allocation process, a balanced distribution was considered in accordance with the land-use statistics. For rare and overrepresented classes, over- and under-sampling was applied respectively to sufficiently represent rare classes of interest and to decrease the workload for widely distributed classes such as grassland.

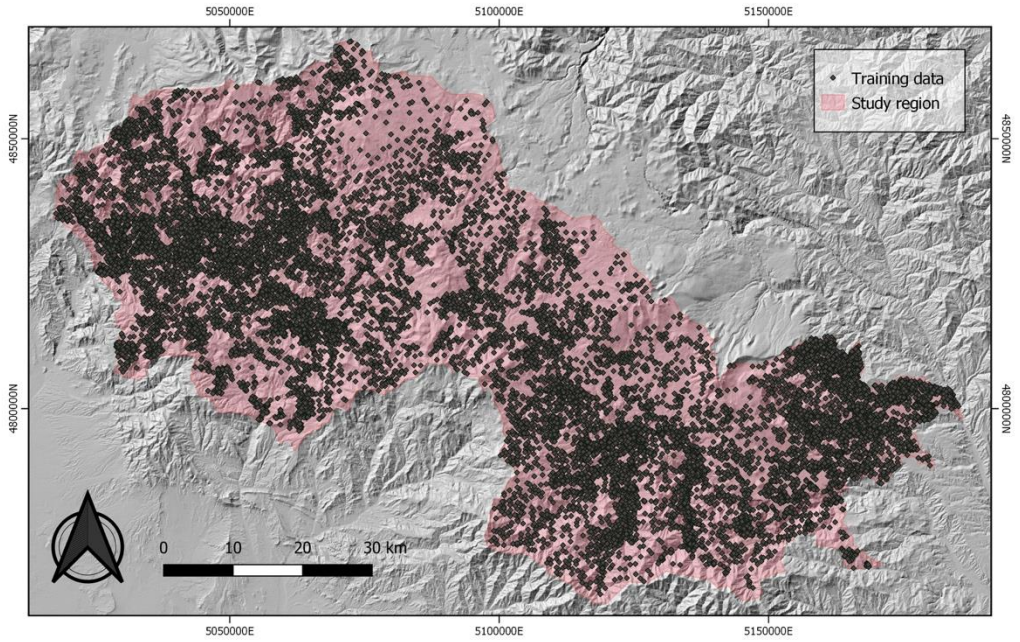


Figure 11 Schematic map showing the distribution of the collected training samples (source: own illustration).

Training data class allocation

The cadaster data provides information about the land use or land cover found in the study region and the associated area proportions, but not regarding their spatially explicit distribution. Due to the implication of the SARS-CoV-2 pandemic, an acquisition of reference data in the field was not possible either. Although recent advances have been made regarding cloud-based remote sensing applications and open access to imagery and platforms that facilitate the interpretation of imagery and the sampling process (e.g. Collect Earth), a lack of knowledge of the situation on the ground and associated reference data may lead to serious misinterpretations (interpreter bias). This was strived to be minimized by a careful review of literature, existing map products, official land balance statistics and consultation with locals and experts.

The class allocation of training data was then entirely based on interpretation of remotely sensed imagery, associated with ancillary data for support (e.g. topographic data) and derived pixel values and time series charts. The assignment of class membership is based on the class typology (section 2.2) and directed by three different interpretation frameworks: (1) Daily images and monthly cloud free composites with full spatial and temporal coverage via the Planet Lab Explorer, (2) GEE framework with multiple information layers, including multi-temporal imagery, time series charts, different quality composites and derivates such as NDVI or bare soil index (BSI)

and (3) Time series visualization framework by He Yin (2018). Figure 12 illustrates a decision tree for the assignment of class membership. A detailed description of the collection framework can be obtained from appendix II-IV.

In order to avoid optimistic bias, only pixels where the membership of a class was known with a very high certainty were also included in the final training sample. The 0.01% share of vine yards in the zone of Sisian-Goris was excluded from the classification. Besides the low expected representation, no area was encountered where a vine yard class membership could be determined with an acceptable certainty. In order to minimize autocorrelation, a minimum distance was considered for the majority of points. However, this was partially violated for rare classes and target classes due to applied over-sampling to ensure satisfactory representation of the spectral variations. Olofsson et al. (2014) stress the importance of a basic visual assessment of the map in order to evaluate the suitability of the map for the given applications before proceeding to a more detailed assessment. This was done repeatedly by adjusting the training data and removing obvious errors in an iterative manner to refine the map before the collection of validation data and the analysis of the reference classification for the accuracy assessment.

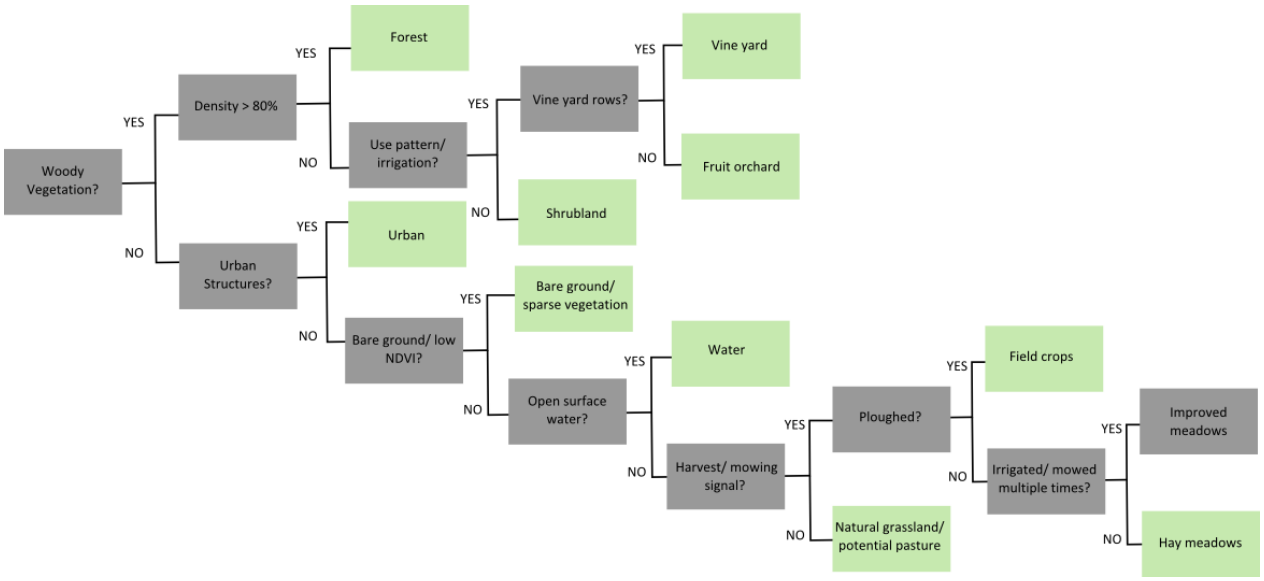


Figure 12 Decision tree for the assignment of class membership based on specific criteria. The final LULC classes are displayed in green (source: own illustration).

Validation data collection

The quality of the predicted thematic maps is determined with an accuracy assessment using validation data. The sampling of validation data was largely based on the "good practice" recommendations for accuracy assessments from Olofsson et al. (2014). The methodology can be separated into the three major components: the sampling design and response design which will be described in this section and the accuracy analysis, which will be described in the next section. Ideally, the validation data should be ground truth and statistically independent from the training data. Following the "good practice" recommendations, this section also outlines the adaptations of the protocols for the objective of this study and reports problematic cases and associated deviations.

Validation data sampling

The allocation of points for probabilistic sampling can be achieved with different approaches such as the use of clusters or strata as well as the use of a simple-random versus systematic sampling There are certain trade-offs between different approaches regarding specific accuracy objectives and desirable sampling design criteria. Common accuracy objectives are the estimation of overall-accuracy, user's accuracy and producer's accuracy or error of commission and error of omission respectively, as well as area of each class. Desirable sampling design criteria include ease and practicality of implementation, representative spatial distribution across the ROI, low variance of the accuracy and area estimates, meaningful sample sizes for rare classes and ease of applying changes in any step of the sampling design.

In this study, a stratified simple-random sampling approach was chosen for the allocation of validation points. Stratification is defined as the partitioning of the ROI into strata where each assessment unit (i.e. pixel) is assigned to a single stratum. Stratification can be applied for two cases: 1) To separate a ROI with pronounced differences of the environment or management systems for better classification results; and 2) To separate a ROI based on the classes obtained from the predicted map in order to report the results (e.g. accuracy and area estimates) per class. Other reasons include the urge to map and report accuracies and class distribution for a specific region (i.e. administrative district) and to stay within quota limits of applications or platforms. For map validation, a stratified sampling approach addresses the key objective of this study to

accurately estimate class-specific accuracies and areas. For this study a stratification was applied on two levels: (1) the level of the sub-regions of Vayots Dzor and Sisian-Goris due to a strong environmental gradient between the two regions and more importantly due to GEE quota limitations for the training sample size; and (2) the level of the LULC classes derived from the predicted map with a stratum for each class for the sampling of validation data.

The choice of sample allocation proportions for the strata is an essential step of the sampling protocol, because different allocations favor different estimation objectives. An allocation of equal sample sizes to each stratum favors user's accuracy over producer's and overall accuracy, while proportional allocation (i.e., the sample size allocation to each stratum is proportional to the area of each stratum) generally lower standard errors of overall and producer's accuracy estimates. Olofsson et al. (2014) therefore advocates a compromise between favoring user's versus producer's and overall accuracies. For this study, the allocation was slightly shifted away from a proportional allocation by increasing the sample size in the rarer classes albeit paying attention not to increase the size to the point where the sample allocation is equal.

The determination of the sample sizes for the allocation of points for each stratum can be derived by calculating the anticipated standard errors for various sample sizes and allocations. Cochran (1977),provide a formula to estimate minimum sample sizes for stratified random sampling (Eq. 2), where N is the number of units in the ROI, $S(\hat{O})$ is the standard error of the desirable estimate of overall accuracy, W_i is the mapped proportion of area of class i, and S_i is the standard deviation of stratum *i* where: $S_i = \sqrt{U_i(1-U_i)}$ (Cochran, 1977). Because N is typically large (e.g., over 11 million pixels in this study), the second term in the denominator of equation 2 can be ignored. The FAO provides an excel spreadsheet where the minimum sample size can be estimated based on equation 2 (FAO, 2016). The spreadsheet was adjusted to the set of class strata of each of the two sub-regions (11 classes in Vayots Dzor and 10 classes in Sisian-Goris). A target standard error of 0.01 was set for the overall accuracy. The mapped proportion of area for the classes (W_i) were obtained from the initially predicted map.

$$n = \frac{(\sum W_i S_i)^2}{[S(\hat{O})]^2 + (\frac{1}{N})\sum {W_i S_i}^2} \approx (\frac{\sum W_i S_i}{S(\hat{O})})^2 \qquad \textit{(Eq. 2)}$$

A literature review was conducted to identify common user's accuracy ranges for the classes of this study. Appendix V shows the acquired user's accuracy estimates of different publications. Although it is difficult to compare the resulting estimators from different studies due to the variety of applied methodologies and data sources, the accuracy estimate ranges may inform about class specific accuracies. The overview revealed that some classes such as forest cover, open water and actively used cropland generally exhibit higher user's accuracies. Other classes such as hay meadows, grassland or shrubland and show lower user's and producer's accuracies. Considering this, expected class-specific user's accuracies were estimated with three levels of optimism. Overall and class-specific sample sizes were calculated for the different class specific user's accuracies. The results are presented in appendix VI for Vayots Dzor. However, this reflects a proportional sample allocation and was modified in order to balance different accuracy objectives as described previously.

For the sample allocation, a focus was put on the agricultural target classes. Although the proportional minimum sample size estimates are small given their little distribution, these classes in particular require larger sample size due to lower anticipated accuracies and due to the fact that they are of specific interest. In turn, for classes with higher anticipated user's accuracies and classes of lower interest, smaller sample sizes are sufficient for the aim of this study. In order to inform about the sample size allocated for each class of interest the variance estimator for user's accuracy was applied (see Section 2.8, Eq. 8). Thus, a class-specific sample size can be identified that is needed to achieve certain standard errors for the assumed user's accuracy for that class. Accordingly, the allocation of samples per class strata can be adjusted until a satisfactory precision of the estimate is reached. It is relevant to mention that insufficient sampling does not result in biased estimators of accuracy and area but imprecise estimators with a higher standard error.

Following this procedure, the sample sizes per class have been determined on the basis of the moderately optimistic assumed user's accuracies (Alloc2, appendix VI), considering primarily the work load and time restrictions in respect to a larger sample allocation. For the by-product classes forest, bare, water and urban sample sizes between 50 and 100 were chosen. For shrubland, a sample size of 200 was chosen due to the larger predicted proportion in Vayots Dzor. For the agricultural target classes, sample sizes of 200 were considered to be sufficient. An

exemption was made for improved meadows with only 150 samples due to the low distribution. For the grassland class, a sample size below 1000 points was chosen due to time constraints.

The table of appendix VII shows the advocated allocations per class strata, the assumed user's accuracies, and the associated standard errors. The determined sample sizes result in acceptable standard errors for the agricultural target classes. In the case of grassland, where the assumed user's accuracy was increased to 90% to reduce the sample size, the standard error didn't increase drastically for smaller assumed user's accuracies and was around 1.57% for an assumed user's accuracy of 70%. For the classes of minor interest such as urban or bare, the expected standard errors of around five percent for the assumed user's accuracy still allow meaningful estimates with a moderate precision.

The sampling approach in GEE is based on a random sample with a minimum distance filter. The exact number of randomly sampled points cannot be directly controlled because close points are only erased subsequently after the allocation (i.e. the distance reduced number of points cannot be controlled directly). Further, since the strata are based on the classes of the map, it is likely that a proportion of pixels are false positive (i.e. a pixel that is classified as grassland but is actually shrubland) and therefore do not account for the validation sample of that class. Hence, an a-priori unknown proportion of all the pixels allocated in each stratum will account for other classes. In particular for rare classes, where sampled pixels are likely to be close and erased or classified as false positive, a satisfactory sample size can be hindered. In order to account for this, the number of allocated points in each stratum has been increased. Since the pixels classified as false positive in one strata account for the validation sample of another, the final sample sizes are likely to be lower for some classes and higher for other classes than initially aimed for. Although this circumstance is inconvenient for a sampling of a specified number of points, GEE provides an option within the applied sampling approach to iteratively repeat the process until a desired sample size in each class is reached.

Validation data class allocation

The sampled points need to be assigned the correct class membership (flagging)as part of the response protocol. The decision of class membership of each point needs to be based on reference data. Further the reference classification should be of higher quality than the classification of the produced map. This can be achieved by using a reference source of superior

quality, or, if using the same reference source, by using a more precise process to create the reference classification (Olofsson *et al.*, 2014). As detailed earlier, the acquisition of reference data faced serious limitations within the scope of this study due to a lack of detailed, spatially explicit archived data and the constraints for an assessment in the field. The reference data sources for the class membership agreement were thus the same as in the collection of training data and based on the same process as outlined for training data. It was strived for a more accurate creation the reference classification by investing more time in the agreement of reference class membership and in the response design and flagging process in general.

Both positively and false positively classified pixels were labeled with the correct reference class membership by placing a point feature with a specific class attribute on the randomly sampled pixels within GEE. During the labeling protocol, both easier and more difficult situations were encountered in respect to determining the reference class membership of the pixel. Some of these cases are described with a short interpretation protocol in appendix VIII-X for transparency. It must be noted that a small portion of the allocated points were not labelled because they were sampled outside the strata of interest. In order to avoid double accounting of these pixels and the associated impact on equal allocation probabilities, these samples have been excluded. Table 3 shows the final allocation of validation points for each strata and the associated land cover proportions from the land balance.

As explained, the finally allocated sample size deviates from the aimed sizes to a certain extent. For some classes like shrubland, hay meadows and fruit orchards, the final sample sizes are larger than aimed for. This is favorable in respect to the precision of the accuracy and area estimates. Accordingly, for some classes less samples could be flagged and the resulting sample size is smaller than initially desired. Therefore, the precision of the estimators is reduced for classes like vine yards, field crops or improved meadows. Due to time constraints, a complementary sampling of validation points for these classes was not applied. The described sampling and response design have been applied for both regions with minor adjustments. In the agroecological zone of Sisian-Goris, there was one class less and the proportions of the validation data are slightly different for both zones. However, problematic pixels encountered during the flagging process have been treated consistently and for both regions, a stratified random sampling was implemented with equal requirements regarding basic accuracy assessment

objectives and the desirable sampling design criteria. The spatial distribution of the final validation samples is shown in figure 13.

Table 3 LULC proportion as derived from the government land balance and proportions of validation samples for both zones of the study region.

	LULC proportions (land balance)				Validation samples (pixel count)			
LULC Class	*Vayots Dzor*		*Sisian-Goris*		*Vayots Dzor*		*Sisian-Goris*	
	Area (ha)	% of total	Area (ha)	% of total	Area (pixel)	% of total	Area (pixel)	% of total
Grassland	119005,51	50,83	119375,38	52,58	851	34,64	812	36,22
Hay meadows	51781,26	22,12	33198,18	14,62	243	9,89	292	13,02
Improved meadows	8630,21	3,69	5533,03	2,44	139	5,66	126	5,62
Field crops	19259,40	8,23	32266,84	14,21	169	6,88	253	11,28
Vine yards	893,30	0,38	29,94	0,01	105	4,27	n.d.	n.d.
Fruit orchards	1011,10	0,43	423,92	0,19	284	11,56	132	5,89
Shrub land	7544,15	3,22	10012,69	4,41	332	13,51	275	12,27
Forest/ Woodland	14697,75	6,28	17943,82	7,90	80	3,26	101	4,50
Sparse vegetation/ bare	2346,00	1,00	303,44	0,13	98	3,99	92	4,10
Water	3297,43	1,41	1925,86	0,85	50	2,04	68	3,03
Urban/ Build up	5655,40	2,42	6012,41	2,65	106	4,31	91	4,06
Total	234121,50		227025,52		2457		2242	

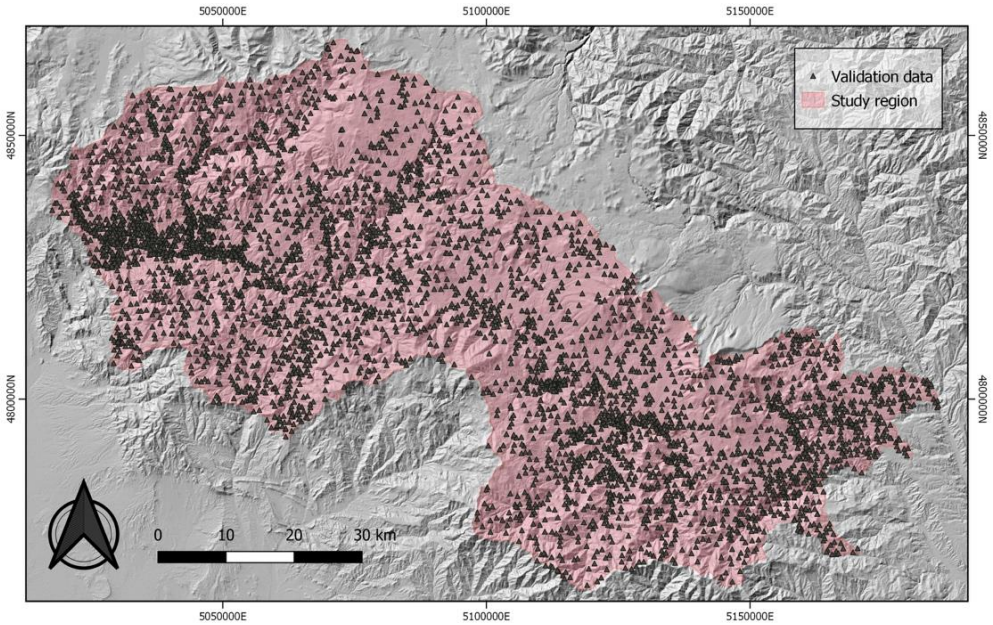


Figure 13 Schematic map showing the distribution of the collected validation samples (source: own illustration).

In order to account for autocorrelation within the validation data a minimum distance filter was set to 500 m for each class stratum. Two exemptions have been made for improved meadows and vine yards with a minimum distance of 250 m only to increase the sample size of the allocated points within these rare classes. In order to assess the autocorrelation for the validation data, Moran's I was calculated using the Moran function implemented in the R raster package (Hijmans *et al.*, 2020). The calculation was run for the predicted maps used for the stratification of the validation data sampling of both sub-regions.

For the calculation of Moran's I, the differences among all possible pixel pairs are assessed. For the whole ROI (6 Mio pixel in each sub-region) this is not feasible regarding the computation power. Moran's I can be therefore calculated for multiple sub-sets of the classified map in different landscapes. This is also favorable regarding major variations between landscapes and the implications for resulting autocorrelation estimates. In order to be representative, a minimum size for the sub-set maps of 30km x 30km was applied. The classified map in appendix XI illustrates seven different landscapes considered for the calculation of Moran's I which are marked with points. Natural and semi-natural land-cover such as grassland and shrubland which are dominant over large areas and clusters of agricultural land use are likely to exhibit spatial autocorrelation. It is of interest to assess the spatial autocorrelation in the map in order to report given values for the validation data and possibly counteract spatial autocorrelation by increasing the minimum distance for the validation samples.

For each of the sub-set classifications, the associated values of Moran's I were calculated with a scale of minimum distance values and the resulting curves were plotted. A distance dependent Moran's I range has been constructed based on the curves of the different sub-sets. The chosen minimum distance thresholds of the validation sampling have been then compared with both the range and specific curves of the subset.

An important last step of the response design is the specification of rules for determining agreement between the obtained map and the reference classification. In the simple case of this study a single label is used for both the map and the reference classification. If these labels agree for a certain assessment unit the predicted map class is correct. If the labels disagree there is a misclassification characterized by the type of confusion defined by both labels. Olofsson et al.

(2014) stress the importance of uncertainty affecting the validation data regarding a precise definition of agreement. Two potential error sources that could lead to a pronounced bias in the accuracy and area estimates are: (1) the uncertainty in respect to spatial co-registration of the map and reference location, and (2) the uncertainty associated with the interpretation of the reference data. In the case of this study, the reference data was obtained from readily co-registered image sources. No data was acquired in the field where the reference is likely to be affected by spatial co-registration for example due to a low quality of the GPS position. The interpreter uncertainty is more relevant in this study. In particular interpreter bias that is defined as an error in the assignment of the reference class to the spatial unit could lead to biased accuracy and area estimates. As mentioned earlier, this is mainly due to the lack of reference data and the resulting approach of simply interpreting imagery in order to derive "ground truth". The ideal would be if the reference classification is based on reference data where the samples represent the exact ground truth. However, in the case of this study the ground truth is constructed from a comprehensive interpretation and therefore likely to be affected by interpreter bias.

2.8.Accuracy and area estimates

The measures to derive accuracy and area estimates of the mapped classes from the map and reference classification are part of the analysis protocol. In the case of this study, there are two essential objectives of the analysis: (1) accuracy assessment of the LULC classification, and (2) estimation of the area of each LULC class. The error matrix (also known as contingency table) is the central tool to assess the accuracy and estimate the class areas of the predicted map. It is a simple cross-tabulation of the class labels of the predicted map classification against the reference classification and the class labels of the validation samples. The major diagonal of the error matrix features the correctly classified samples whereas the off-diagonal elements show the errors of omission and commission (Olofsson *et al.*, 2014). Appendix XII shows an example of a basic 4x4 error matrix with four different classes.

The rows of the error matrix display the labels of the predicted map classification whereas the columns show the labels of the validation samples from the reference classification. Within the matrix, p_{ij} represents the area or pixel count proportion of the population that has the map

class *i* and the reference class *j*, where population is defined as the full region of interest. From a population error matrix with *q* classes, the overall accuracy O is given by equation 3. The user's accuracy U_i is a measure of the proportion of the area correctly mapped as class *i* that has reference class *i* (Eq. 4). Its inverse parameter $(1 - U_i)$ is the error of commission. The producer's accuracy P_j is a measure of the proportion of the area of reference class *j* that is correctly mapped as class *j* (Eq. 5). Its inverse parameter $(1 - P_i)$ is the error of omission.

$$O = \sum_{j=1}^{q} p_{jj} \qquad \textit{(Eq. 4)}$$

$$U_i = p_{ii}/p_{i\cdot} \qquad \textit{(Eq. 5)}$$

$$P_j = p_{jj}/p_{\cdot j} \qquad \textit{(Eq. 5)}$$

Although there are many other measures of accuracy such as the commonly used kappa coefficient of agreement, Olofsson et al. (2014) suggest to use only the three basic accuracy metrics presented above, arguing that kappa is highly correlated with overall accuracy and it would be redundant to report both measures.

To derive the estimates of accuracy and area from the validation sample, Olofsson *et al.*, (2014) advocate for a design-based inference as framework for the derivation estimates. A central principle of the design-based inference is that the specific estimator depends on the applied sampling design which means that different sampling designs require different estimators. First, the cell entries of the error matrix and the derived measures have to be estimated from a sample. Once the error matrix is constructed and p_{ij} is known, the sample-based estimator of p_{ij} $(\hat{p}_{ij})$ can be calculated. The error matrix may also be reported as the sample-based area proportions $\hat{p}_{ij}$ rather than sample counts p_{ij}. The formula to derive $\hat{p}_{ij}$ depends on the implemented sampling design of the validation data. In case of probabilistic designs like the stratified random sampling it is given by equation 6,

$$\hat{p}_{ij} = W_i \frac{n_{ij}}{n_i} \qquad \textit{(Eq. 6)}$$

where W_i is the proportion of area mapped as class *i*. Substituting $\hat{p}_{ij}$ in equations 3-5 results in sample-based estimators of overall, user's and producer's accuracies.

The sampling variability linked with the accuracy estimates should be assessed by calculating and reporting standard errors. The variance of the sample-based estimator of overall accuracy, $\hat{V}(\hat{O})$, is given by equation 7. The variance of the sample-based estimator of user's accuracy, $\hat{V}(\hat{U}_i)$, is given by equation 8 and the variance of the sample-based estimator of producer's accuracy is given by equation 9, where $\hat{N}_{\cdot j} = \sum_{i=1}^{q} \frac{N_{i\cdot}}{n_{i\cdot}} n_{ij}$ is the estimated marginal total number of pixels of reference class *j*. $N_{j\cdot}$ is the marginal total of map class *j* and $n_{j\cdot}$ is the total number of sample units in map class *j*. The square root of the variance estimators results in the standard errors. The variance estimators are based on an assessment unit of one pixel and each pixel has a single map and reference classification label. The square root of the variance estimators results in the standard errors (Olofsson *et al.*, 2014).

$$\hat{V}(\hat{O}) = \sum_{i=1}^{q} W_i^2 \, \hat{U}_i(1 - \hat{U}_i)/(n_{i\cdot} - 1) \qquad \textit{(Eq. 7)}$$

$$\hat{V}(\hat{U}_i) = \hat{U}_i(1 - \hat{U}_i)/(n_{i\cdot} - 1) \qquad \textit{(Eq. 8)}$$

$$\hat{V}(\hat{P}_j) = \frac{1}{\hat{N}_{\cdot j}^2}\left[\frac{N_{j\cdot}^2(1 - \hat{P}_j)^2 \hat{U}_j(1 - \hat{U}_j)}{n_{j\cdot} - 1} + \frac{\hat{p}_j^2 \sum_{i \neq j}^{p} N_{i\cdot}^2 \frac{n_{ij}}{n_{i\cdot}}\left(1 - \frac{n_{ij}}{n_{i\cdot}}\right)}{n_{i\cdot} - 1}\right] \qquad \textit{(Eq. 9)}$$

Apart from the accuracy estimates, the error matrix also serves as foundation for the area estimates of each class. The class specific area may be derived from the error matrix in two different ways. There is interest in estimating the proportion of area of class j=*k*. The row and column totals are the sum of the p_{ij} values in the corresponding rows and columns. Consequently,

the row total $p_{k\cdot}$ is the proportion of area mapped as class *k* and the column total $p_{\cdot k}$ represents the area proportion of class *k* as specified by the reference classification. Because the reference classification is generally less biased than the predicted map (i.e. the bias associated with reference data error is smaller than the bias associated with map classification error) it is recommended, that the estimation of class specific area is based on $p_{\cdot k}$. For the stratified random sampling, the widely applied area estimator $\hat{p}_{\cdot k}$ is recommended by Olofsson et al. (2014), its formula is given in equation 10. The standard error of the stratified area estimator $\hat{p}_{\cdot k}$ is given by equation 11, where n_{ik} is the sample count of cell element i of class k in the error matrix, W_i is the area proportion of map class *i*, $\hat{p}_{ik} = W_i \frac{n_{ik}}{n_{i\cdot}}$ and the summation is over the *q* classes.

$$\hat{p}_{\cdot k} = \sum_{i=1}^{q} W_i \frac{n_{ik}}{n_{i\cdot}} \qquad \textit{(Eq. 10)}$$

$$S(\hat{p}_{\cdot k}) = \sqrt{\sum_i W_i^2 \frac{\frac{n_{ik}}{n_{i\cdot}}(1-\frac{n_{ik}}{n_{i\cdot}})}{n_i - 1}} = \sqrt{\sum_i \frac{W_i \hat{p}_{ik} - \hat{p}_{ik}^2}{n_{i\cdot} - 1}} \qquad \textit{(Eq. 11)}$$

3. Results

In this section, the results of the LULC classification are presented and illustrated. Due to the stratification into agroecological zones, the results are displayed separately for Vayots Dzor and Sisian-Goris. First, the LULC class spectra are assessed with descriptive statistics and illustrated to assess class separability and limitations. Then the accuracy and area estimates of the "good choice" classification are provided and the distribution and extent of LULC classes is illustrated with a map. Major misclassification is assessed and described with a focus on the agricultural target classes. Maps for two simple applications of the LULC data are displayed. Subsequently, the baseline classification is compared to the different dataset trials to assess the added value of using multisource data, illustrated with both tables and bar charts of the user's accuracy. Major differences between the datasets are described and trends of enhancement of the classification are highlighted. Apparent visual differences between the baseline and the "good choice" classification are further illustrated with a map. Finally, the spatial autocorrelation results

for the validation data are displayed to assess the potential impact on the quality of the accuracy and area estimates.

3.1.Spectral class assessment

Class separability is of particular interest to understand the spectral collinearity of the agricultural and semi-natural classes. The knowledge capture was achieved by sampling pixel values of the different bands of the training samples and visualized with box-plots and time series line charts. The knowledge capture is useful to identify input variables which are more likely to effectively differentiate classes as well as those which are prone to fail in providing a distinction. The time series illustrate differences in plant phenology of the classes over a growing season and provide insight which months are more suitable for the discrimination of classes.

Figure 14 shows both charts for NDVI and all agricultural and semi-natural classes in Vayots Dzor with (a) the box- and whisker-plots for the spectral-temporal metrics, and (b) the monthly time series as line chart. The corresponding results for Sisian-Goris are provided in appendix XIII. As expected, the grassland and hay meadow class are spectrally similar across all spectral-temporal statistics with some minor differences. On the one hand, the maximum NDVI is higher for hay meadows, indicating a higher biomass, which is most likely why these areas are used for hay making. On the other hand, the minimum NDVI is slightly lower for the hay meadows which might be associated with the hay cutting. Therefore, the standard deviation is also higher for hay meadows. Improved meadows exhibit high NDVI values across all spectral-temporal metrics. Whereas the maximum NDVI is similar to hay meadows, the minimum, the mean and the quartiles show higher values. The time series charts mirror those observations with higher peak values of the hay meadows in spring compared to grassland and a steadily high NDVI for improved meadows.

The cropland class in Vayots Dzor is spectrally similar to grass land and hay meadows. Besides an exceptionally low standard deviation and low values across all statistics, the time series of the field crops and grassland class are very collinear as well. The NDVI time series of the field crop class in Sisian-Goris shows a similar trend (see appendix XIII). The classes of vine yards, fruit orchards and improved meadows also show a related phenology and mutual spectral signals. The NDVI time series shows that the phenology of improved meadows and fruit orchards are very

similar. This emphasizes the role of the grass signal for these classes. Both are irrigated and characterized by a strong herbaceous signal. The vine yard class has a less pronounced grass signal, likely due to a higher rock content in a drier substrate. The phenology of shrubland is very similar to that of grassland, but slightly higher in all spectral-temporal metrics. In Vayots Dzor, shrubland has lower NDVI values across the temporal statistics than the other woody classes, but a higher standard deviation than vine yards. In general, all woody classes have high NDVIs and it is increasing with canopy volume and density. Accordingly, forests show a distinct spectra-temporal signal in comparison to the other classes.

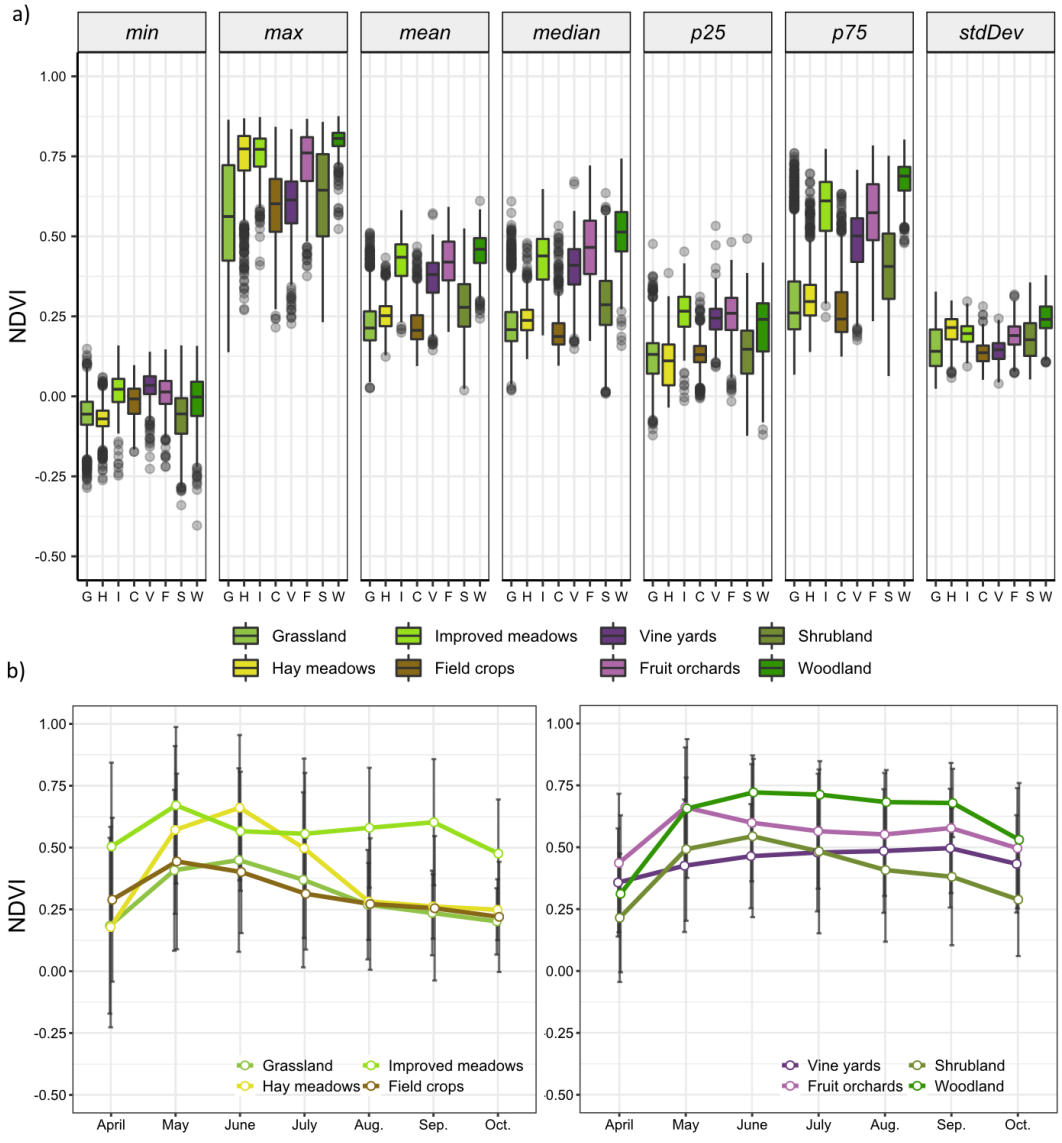


Figure 14 Knowledge capture of the spectral-temporal characteristics of the agricultural and semi-natural LULC classes for Vayots Dzor. (a) the box- and whisker-plots for the spectral-temporal metrics, and (b) the monthly time series as line chart. The outliers in the box-plots are depicted as grey dots. Both graphs were created in R using ggplot2.

The other original optical bands and the optical derivates such as NIR or NDMI show a matching trend (see appendix XIV & XV). The BSI shows an opposite behaviour than the NDVI and is especially pronounced for the field crops due to ploughing and fallow periods (see appendix XVI). The Tasseled Cap metrics show a lower potential for class separability than other indices. For Sisian-Goris, the spectral-temporal similarity of the shrubland and fruit orchard class is more pronounced and fruit orchards are characterized less with the herbaceous signal of the improved meadows (see appendix XIII).

Based on the literature it is clear that the visual spectra have a lower ability to separate narrow classes (Mahdianpari *et al.*, 2019). The box- and whisker plots in appendix XIV show that NIR and SWIR bands show the phenological and spectral differences but less pronounced than some of the indices (NDMI, MSAVI2 in appendix XV). The phenological class characteristics obtained from the time series are also more lucid for the vegetation indices than for the original NIR-band (see appendix XVII).

The SAR data and the derived monthly time series show very similar temporal trends for the different agricultural and semi-natural classes (see figure 15). For the VV composites, only the hay meadows and the woodland class can be unambiguously distinguished from the other classes. The time series of the VH composites shows a corresponding behaviour with even less pronounced differences (see appendix XVII).

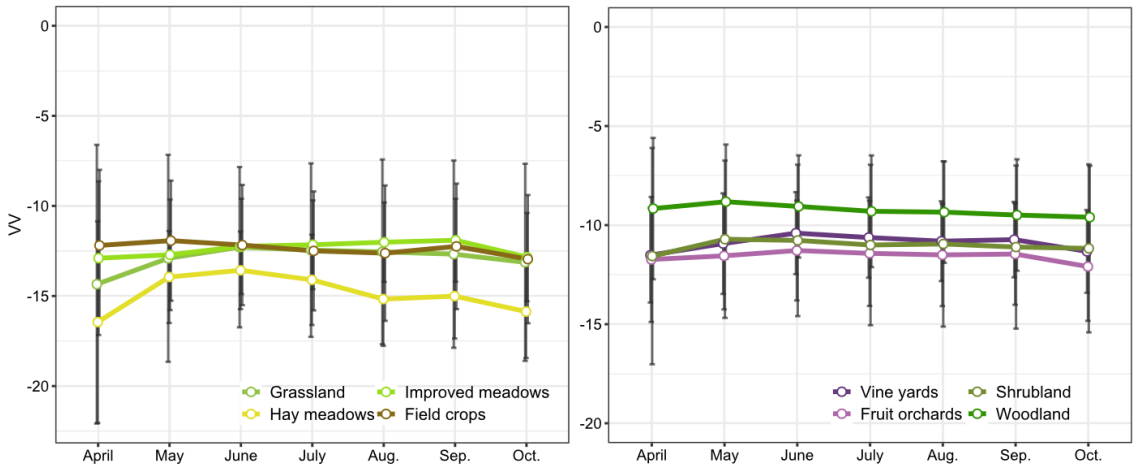


Figure 15 Time series of the VV composites for the agricultural and semi-natural classes (source: own illustration).

In order to inform about the trends related to the topography and learn about the potential effects of including ancillary topographic data, the training data of the whole ROI (Vayots Dzor and Sisian-Goris) was used to sample the hill slope values derived from the digital elevation product. The classes were grouped in: (1) agricultural classes which include hay meadows, improved meadows, field crops, vine yards and fruit orchards; (2) semi-natural classes that include grassland, shrubland, forest and bare surfaces or sparse vegetation; and (3) other classes including the urban and water class. The box- and whisker-plot of figure 16 shows the dependency of the agricultural classes and the topography. While 75% of the agricultural classes have a slope of less than 10%, 75 % of the semi-natural classes have a slope of more than 10%.

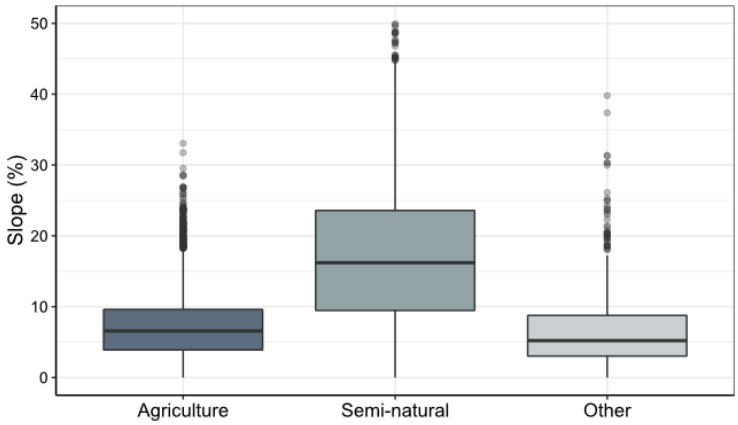


Figure 16 Box- and whisker-plots illustrating the relation of land-use and land cover classes with the hill slope as topography metric (source: own illustration).

3.2. "Good choice" classification

The "good choice" dataset is based on the observations from the classification trials with the multisource data (see section 3.3). The combined datasets delivered the most favourable map prediction in terms of user's accuracies as priority mapping objective. The two sub-regions varied regarding an enhancement of the mapping accuracy when including multisource data. Based on the variations, the good choice dataset was adjusted to each of the sub-regions, because a detailed and accurate representation of the complex LULC was a main priority of this study. Sample count-based error matrices with area adjusted overall, producer's and user's accuracy estimates of the "good choice" classification for both sub-regions are shown in table 4. Vayots Dzor comprises 11 classes, Sisian-Goris only 10 classes due to the exclusion of the vine yard class. An overview of the RF feature importance can be obtained from the bar chart in appendix XVIII.

Table 4 Sample count-based error matrix with area-adjusted accuracy estimates and standard errors of the "good choice" classification. User's accuracy (UA) and producer's accuracy (PA) are shown besides and below the marginal totals respectively with the associated standard errors depicted in blue. Overall accuracy (OA) is shown in pink in the bottom right. Top: the 11 classes of Vayots Dzor. Bottom: the 10 classes of Sisian Goris.

		Reference													
		GL	HM	IM	FC	VY	FO	SL	WL	SP	W	U	∑	UA (%)	±
Map	GL	549	23	6	38	15	10	40	0	1	0	9	691	79,45	1,61
	HM	21	186	5	6	0	0	1	0	0	0	0	219	84,93	2,45
	IM	0	2	97	0	19	18	2	0	0	0	0	138	70,29	3,33
	FC	6	3	8	127	15	3	0	0	1	0	0	163	77,91	2,96
	VY	0	0	3	2	93	6	1	0	0	0	0	105	88,57	2,14
	FO	1	0	62	6	46	159	2	0	0	0	1	277	57,40	3,12
	SL	46	0	7	11	18	44	163	7	4	8	5	313	52,08	3,43
	WL	0	0	1	0	0	2	2	68	0	0	1	74	91,89	3,13
	SP	2	0	0	2	0	0	0	0	65	2	2	73	89,04	3,71
	W	1	0	0	0	0	0	2	1	0	45	0	49	91,84	3,69
	U	1	0	0	5	17	10	0	1	1	1	68	104	65,38	5,16
	∑	627	214	189	197	223	252	213	77	72	56	86	2206		
	PA (%)	95,76	61,33	12,74	18,54	20,99	22,30	61,71	82,67	91,47	23,85	19,90		OA (%)	±
	±	0,02	0,55	1,83	0,51	1,13	1,18	0,34	2,78	0,63	17,17	2,84		76,23	1,27

		GL	HM	IM	FC	FO	SL	WL	SP	W	U	∑	UA (%)	±
Map	GL	647	33	7	10	30	23	0	23	1	9	783	82,63	1,43
	HM	28	243	5	13	2	1	0	0	0	0	292	83,22	2,19
	IM	2	4	105	2	7	6	0	0	0	0	126	83,33	3,00
	FC	11	12	4	223	2	0	0	1	0	0	253	88,14	2,00
	FO	0	1	20	2	97	7	5	0	0	0	132	73,48	2,92
	SL	18	0	14	7	69	152	9	1	0	5	275	55,27	3,46
	WL	0	0	0	0	9	13	79	0	0	0	101	78,22	4,30
	SP	0	0	0	0	0	0	0	73	0	7	80	91,25	2,83
	W	1	0	0	0	1	2	0	0	62	1	67	92,54	3,34
	U	0	0	0	4	12	3	0	3	0	69	91	75,82	4,51
	∑	707	293	155	261	229	207	93	101	63	91	2200		
	PA (%)	96,97	70,33	8,10	84,73	11,39	58,33	91,70	66,54	85,73	29,63		OA (%)	±
	±	0,02	0,31	1,45	0,32	0,49	1,19	1,08	0,54	2,66	5,48		81,64	1,02

The "good choice" classification delivers an overall mapping accuracy of 76.23% (±1.27%) and 81.64% (±1.02%) for Vayots Dzor and Sisian-Goris, respectively. The producer's accuracies (i.e. the probability that a certain land cover of an area on the ground is classified accordingly) vary largely between some of the classes. Semi-natural LULC classes of Vayots Dzor including

grassland, hay meadows, shrubland and woodland as well as sparse vegetation show higher producer's accuracies with 60-90%. In contrast, the agricultural and artificial classes including improved meadows, field crops, vine yards, fruit orchards as well as urban and water show lower producer's accuracies with 15-25%. For Sisian-Goris, similar trends can be observed with some exemptions such as higher producer's accuracies in case of the field crop and water classes. The high errors of omission of the artificial and agricultural classes in Vayots Dzor indicate limitations with a high proportion of samples that have been left out (omitted) and falsely classified.

The grassland class shows a rather small omission error mainly from misclassification as hay meadows and shrubland. Hay meadows are in turn often misclassified as grassland. Improved meadows and fruit orchards are highly collinear resulting in low producer's accuracies for both classes. Improved meadows were most often misclassified as orchards, but less vice versa. The low producer's accuracies of the orchards are mainly due to confusion with the shrubland class and, to a smaller extent, with improved meadows and grassland. Vine yards also show profound mapping limitations and high omission errors due to the confusion with several classes but especially orchards, improved meadows and shrubland. Corresponding to the confusion of grassland as shrubland, the largest share of the shrubland reference samples are omitted as grassland.

The map reliability is expressed by the user's accuracy or the error of commission as its complementary measure and provides information of how often the class represented on the map will actually be present on the ground. The grassland and hay meadow class show an almost equal mapping reliability in both sub-regions with about 80% user's accuracy and mutual class confusion. A small proportion of the map's grassland class are falsely classified hay meadows, field crops, fruit orchards and shrubland. For both sub-regions, the user's accuracy of improved meadows is diminished mainly by falsely classified vine yards and orchards.

The error of commission for field crops in Vayots Dzor results from falsely classified map pixels actually belonging to grassland, vine yards or improved meadows. In Sisian-Goris there is far less confusion of field crops and those classes and rather an overestimation at the cost of grassland and hay meadows. Vine yards in Vayots Dzor only show a minor error of commission. In contrast, fruit orchards in Vayots Dzor are overestimated in particular by pixels that are actually vine yards and improved meadows. In Sisian-Goris, the user's accuracy for fruit orchards is higher

and misclassifications come at the expense of both improved meadows and shrubland. Shrubland shows a low user's accuracy in both sub-regions with considerable overestimation at the cost of grassland, orchards or vine yards. The map reliability of the urban class is also quite low with overestimates at the cost of fruit orchards in Sisian-Goris and fruit orchards and vine yards in Vayots Dzor. In Sisian-Goris, where vine yards were not included as a class, the user's accuracy of improved meadows and fruit orchards is 83.33 and 73.47% respectively as opposed to 70.29 and 57.4% in Vayots Dzor. Table 5 shows the area estimates with the associated standard errors in hectares for both sub-regions.

Table 5 Area estimates of the 11 LULC classes of Vayots Dzor (VD) and the 10 classes of Sisian-Goris (SG) with the associated standard errors in hectares.

	Class	*GL*	*HM*	*IM*	*FC*	*VY*	*FO*	*SL*	*WL*	*SP*	*W*	*U*
VD	Area (ha)	138092	14504	3929	13378	8269	9217	25722	4035	7745	1323	3664
	±	2645	1165	661	1496	1016	976	1715	281	426	313	769
SG	Area (ha)	135928	26293	3516	23075	-	12696	14449	6913	14599	1427	4351
	±	2225	1280	626	842	-	1204	1105	374	1027	208	711

Figure 17 shows small-scale map depictions of different agricultural landscapes within the study region as predicted by the "good choice" RF model. The area around Yeghegnadzor exhibits a large proportion of vine yards and fruit orchards (a & b). Based on a visual screening of the "good choice" map and Google Earth data in GEE, the impression emerges that the classification is more accurate the larger the plots. The vine yard plots for example which are commercially used seem to be more accurate than smaller plots within and around the communities. Due to the small-scale character within the villages, there is considerable confusion of mixed urban pixels with vine yards as in case of Maslishka village (see figure 17; b). The community of Rind in turn exhibits larger vine yard plots which could be mapped more accurate (see figure 17; west of a). Besides the plot sizes, successional cultivation stages may be responsible for difficulties regarding the mapping accuracies. Whereas fully developed plots of vine yards or orchards are mapped increasingly accurate, less developed plots with smaller plants (i.e. new plantations) show signs of confusion with improved meadows, field crops, bare or urban. Finally, the accuracy results suggest that the fruit orchards within the home gardens of the villages are slightly overestimated which can be observed across the study region.

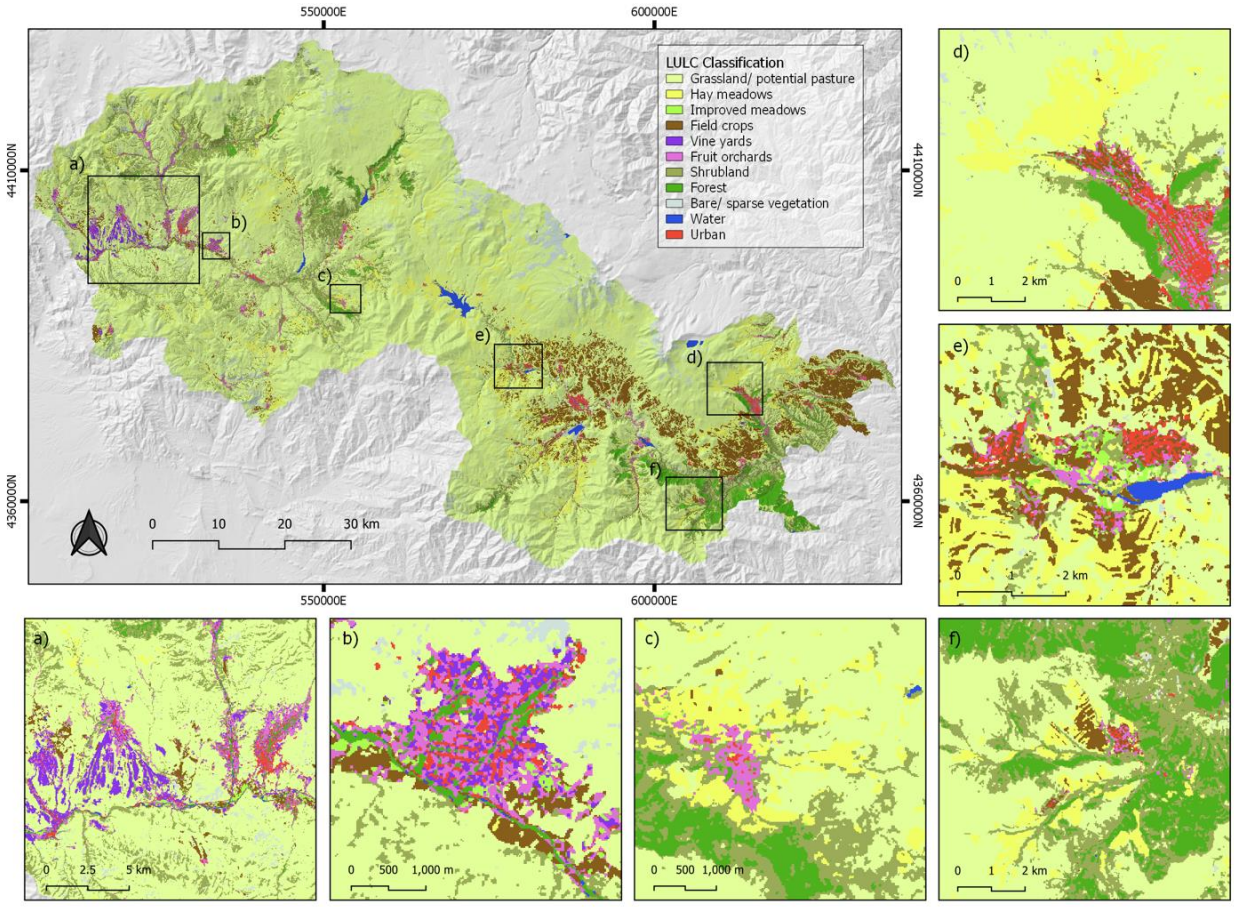


Figure 17 Map with small-scale depictions of different agricultural landscapes as predicted by the "good choice" RF model for both regions Vayots Dzor (a-c) and Sisian-Goris (d-f) (source: own illustration).

Subalpine landscapes with a high heterogeneity such as the area around Artavan village (figure 17; c) show a moderately well distinction between shrubland, forest, fruit orchards, hay meadows and grassland. With increasing altitude, the field crops class becomes rare and limited to single small plots within villages (see figure 17; c). There is evidence that the hay meadows are mapped with a high accuracy (i.e. 84% in Vayots Dzor). However, with increasing altitude the mowing signal seems to be less pronounced and the influence of livestock grazing activities appears to be responsible for some confusion of grassland as hay meadows and vice versa. Hay meadows within landscapes dominated by shrubs show less signs of confusion (see figure 17; c & f). Other more elevated areas with a high herbaceous net primary production such as the subalpine meadows on the slopes below Lake Sevlich also show less signs of confusion with grassland (see figure 17; d). Areas characterized by abundant smallholder farming and a high heterogeneity such as the landscape around Angeghakot give the impression of a successful

distinction between the various classes (see figure 17; e). However, with a closer look it seems that there is a certain degree of class confusion, in particular within very heterogenous, small-scale managed areas. The field crop class seems to be mapped predominantly well with some exemptions where the separability between the crops and hay meadows appears to be distorted. Despite of the topographic data, some areas with very steep slope such as the canyon close to Tatev (see figure 17; f) show signs of confusion with the water and urban class.

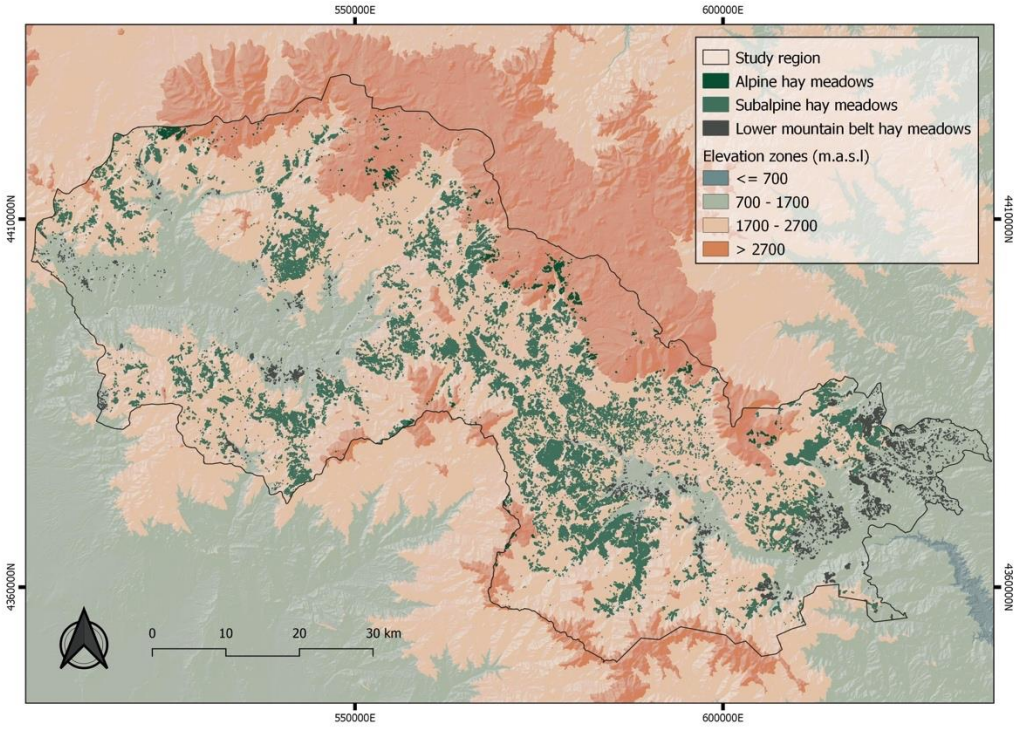


Figure 18 Thematic map showing the distribution of extensively used hay making areas across the different elevation zones in the Armenian Highlands (source: own illustration).

The two thematic maps in figure 18 and 19 illustrate possible applications of the produced LULC data that can be considered as relevant in the given context of land use related challenges in Armenia. The monitoring of hay making as grassland use and the assessment of spatial dynamics such as increasing use close to settlements at the expense of hay making areas in more remote subalpine and alpine zones can be a of significant value regarding the adverse environmental effects of unbalanced pasture loads (see figure 18).

Wildlife management and conservation via the designation of protected areas is of particular relevance within the study region. The high typological detail of land use in the proposed mapping approach exhibits potential for the management of both conservation areas and wildlife corridors (see figure 19). This may include a spatial analysis of human activities affecting wildlife or an assessment of the probability of evolving challenges within the designation process of protected areas due to major land-use conflicts.

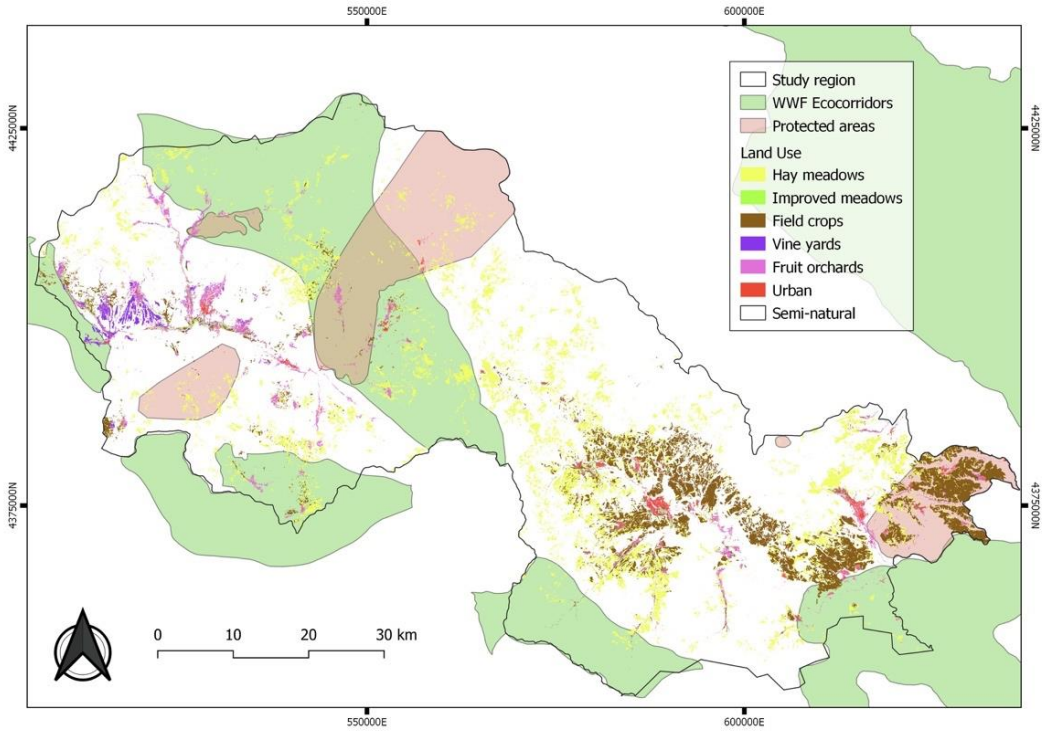


Figure 19 Thematic map of land use within and around major protected areas (red) and wildlife corridors (green) (source: own illustration).

3.3. Dataset classification trials

The baseline dataset that includes only optical data already yielded acceptable mapping accuracies. With 77.15% and 81.51% for Vayots Dzor and Sisian-Goris respectively, the baseline classification yielded similar or slightly higher overall accuracies compared to the “good choice” classification. However the “good choice” prediction shows a higher user’s accuracy , highlighting the trade-off between producer’s and user’s accuracies. The sample count-based error matrices of the baseline classification with area adjusted accuracy estimates and associated standard errors for both sub-regions are provided in appendix XIX & XX.

Table 6 shows the accuracy estimates of the RF predictions based on the different input datasets for Vayots Dzor, table 7 for Sisian-Goris (see section 2.5 for dataset construction). The inclusion of the topography metrics led to a slight increase in the overall accuracy compared to the baseline dataset. The producer's accuracy increased more pronounced for fruit orchards and sparse vegetation whereas it also decreased by 4% in case of hay meadows. The user's accuracy increased more notably in case of hay meadows, field crops and fruit orchards and decreased by a small extent for improved meadows and sparse vegetation. This beneficial effect is however only the case for Vayots Dzor. The inclusion of topography metrics in Sisian-Goris led to a slight deterioration of the map prediction across the different accuracy estimates (table 7).

Table 6 Accuracy estimates of the RF models with the different input datasets (1-7) for Vayots Dzor.

Datasets	OA (%)	User's Accuracy (%)										
		LULC Classes										
		GL	HM	IM	FC	VY	FO	SL	WL	SP	W	U
1: Baseline	77,15	81,48	81,74	66,67	71,78	87,62	51,99	52,72	82,43	80,82	91,84	55,77
2: 1 + Topo	77,26	81,04	83,56	63,04	74,23	87,62	54,87	55,91	82,43	76,71	93,88	56,73
3: 1 + S1 time series	76,63	81,62	76,71	66,67	66,26	82,86	40,43	55,27	85,14	75,34	87,76	68,27
4a: 1 + S2 texture	75,96	80,75	83,11	67,39	73,62	85,71	52,71	46,96	81,08	80,82	95,92	61,54
4b: 1 + S1 texture	76,52	80,90	81,74	67,39	71,78	84,76	46,93	53,99	82,43	73,97	89,80	53,85
5: 1 + VIIRS	77,48	81,77	82,65	63,77	75,46	86,67	51,62	53,35	82,43	80,82	91,84	60,58
6a: 1 + Indice Set 01	76,10	79,74	83,56	63,77	74,23	84,76	54,87	51,12	90,54	87,67	83,67	60,58
6b: 1 + Indice Set 02	76,13	80,61	81,74	65,94	75,46	87,62	51,26	49,20	83,78	80,82	91,84	51,92
6c: 1 + Indice Set 03	75,76	79,59	79,45	66,67	76,07	87,62	57,04	51,76	93,24	83,56	89,80	56,73
7: "Good Choice"	76,23	79,45	84,93	70,29	77,91	88,57	57,40	52,08	91,89	89,04	91,84	65,38
	OA (%)	Producer's Accuracy (%)										
		LULC Classes										
		GL	HM	IM	FC	VY	FO	SL	WL	SP	W	U
1: Baseline	77,15	94,28	66,59	14,40	20,19	22,51	24,73	63,59	81,41	84,39	26,81	17,59
2: 1 + Topo	77,26	94,33	62,52	12,97	20,82	20,52	28,43	63,36	83,06	88,93	27,80	17,01
3: 1 + S1 time series	76,63	93,09	64,71	13,73	19,91	22,33	19,32	64,20	72,74	75,20	47,26	33,51
4a: 1 + S2 texture	75,96	94,24	64,91	13,46	21,46	22,64	23,48	62,96	77,30	77,71	23,00	20,98
4b: 1 + S1 texture	76,52	93,93	63,13	12,65	21,49	23,22	24,54	61,39	83,42	79,48	26,35	18,97
5: 1 + VIIRS	77,48	94,25	66,09	12,43	21,60	22,53	28,22	63,57	81,23	84,41	27,73	26,90
6a: 1 + Indice Set 01	76,10	95,14	59,70	12,66	20,69	17,73	20,94	63,10	74,65	84,31	32,68	19,95
6b: 1 + Indice Set 02	76,13	94,34	66,51	12,82	19,88	22,02	21,54	61,88	73,68	85,48	29,65	14,91
6c: 1 + Indice Set 03	75,76	95,48	59,27	12,11	21,10	19,83	20,93	66,04	73,03	87,59	24,47	15,28
7: "Good Choice"	76,23	95,76	61,33	12,74	18,54	20,99	22,30	61,71	82,67	91,47	23,85	19,90

The inclusion of the monthly SAR time series in Vayots Dzor deteriorated the accuracy of the RF classification slightly compared to the baseline map prediction. The overall accuracy decreased marginally. The user's accuracy decreased more markedly for fruit orchards, vine yards, hay meadows and field crops but increased for urban fabric, shrubland and woodland. This beneficial effect of including SAR data could also be observed in Sisian-Goris. In contrast to Vayots Dzor, the overall accuracy slightly increased as well (see table 7). For fruit orchards, the user's accuracy improved notably by around 7% and for urban structures by more than 10%.

In Vayots Dzor, both optical and SAR texture metrics diminished the overall accuracy as opposed to the baseline. For optical texture metrics this effect was stronger. In contrast, it can be noted that for the agricultural target classes the mean user's accuracies marginally improved in both regions with the optical texture metrics. In Sisian-Goris, optical texture metrics showed a slight enhancement of overall, user's and producer's accuracies. Contrarily, the SAR texture metrics generally reduced the accuracy metrics with some exemptions as in case of the user's accuracy of shrubland or woodland where the inclusion lead to a small increase.

The inclusion of nighttime lights in Vayots Dzor barely increased the overall accuracy by half a percent. However, the user's accuracy increased as expected for the urban class and, more surprisingly, also for field crops due to a reduced confusion with grassland. In Sisian-Goris, the nighttime lights didn't improve the different accuracy estimates with the exemption of a notable improvement of the user's accuracy for the urban class.

Adding normalized difference indices and spectral transformations generally enhanced the user's accuracies of some of the classes, but slightly lowered overall accuracy in (exemption for index set 1 in Sisian-Goris). The lower overall accuracy results mainly from deteriorated producer's accuracies. Single vegetation indices favoured the user's accuracy of only some classes while diminishing the producer's accuracies of most of the classes. A stronger improvement of user's accuracies for single indices could be observed for the MNDWI and the field crop class as well as the EVI and BSI with the fruit orchard class (results not shown). In contrast, sets of indices provided a broader enhancement of the user's accuracies in both sub-regions.

Table 7 Accuracy estimates of the RF models with the different input datasets (1-7) for Sisian-Goris.

Datasets	OA (%)	User's Accuracy (%)									
		LULC Classes									
		GL	HM	IM	FC	FO	SL	WL	SP	W	U
1: Baseline	81,51	83,78	80,14	69,84	85,38	66,67	56,73	78,22	82,50	89,55	63,74
2: 1 + Topo	80,63	82,50	79,79	65,87	87,35	65,15	56,73	76,24	78,75	94,03	64,84
3: 1 + S1 time series	81,70	83,91	79,45	64,29	86,17	73,48	58,91	79,21	82,50	86,57	74,73
4a: 1 + S2 texture	81,92	84,55	77,40	71,43	87,35	68,94	56,36	78,22	81,25	89,55	65,22
4b: 1 + S1 texture	81,45	83,78	78,42	64,29	86,56	65,15	59,27	79,21	77,50	89,55	63,74
5: 1 + VIIRS	81,37	84,04	77,40	68,25	86,17	64,39	55,27	79,21	80,00	89,55	68,13
6a: 1 + Indice Set 01	81,61	83,78	79,45	82,54	86,17	68,18	55,64	78,22	86,25	89,55	67,03
6b: 1 + Indice Set 02	81,12	82,89	78,77	82,54	86,96	68,18	55,64	80,20	86,25	89,55	68,13
6c: 1 + Indice Set 03	80,40	81,61	79,45	79,37	87,35	71,21	54,91	78,22%	87,50	95,52	74,73
7: ”Good Choice”	81,64	82,63	83,22	83,33	88,14	73,48	55,27	78,22	91,25	92,54	75,82
	OA (%)	Producer's Accuracy (%)									
		LULC Classes									
		GL	HM	IM	FC	FO	SL	WL	SP	W	U
1: Baseline	81,51	95,55	69,75	8,15	74,83	14,69	55,23	91,76	71,60	72,21	24,17
2: 1 + Topo	80,63	95,79	67,34	7,52	76,34	12,08	53,63	92,35	60,40	73,50	32,10
3: 1 + S1 time series	81,70	95,63	69,56	7,54	74,45	14,19	61,26	91,11	67,02	100,00	41,32
4a: 1 + S2 texture	81,92	95,36	71,06	7,59	77,87	14,04	56,41	92,43	71,64	82,45	25,91
4b: 1 + S1 texture	81,45	95,35	68,69	7,23	76,44	13,45	57,74	92,27	67,06	85,19	25,05
5: 1 + VIIRS	81,37	95,36	70,43	7,59	75,09	12,85	55,80	90,93	72,66	82,25	24,62
6a: 1 + Indice Set 01	81,61	95,67	70,11	9,05	82,12	12,41	55,27	91,36	72,20	74,12	25,40
6b: 1 + Indice Set 02	81,12	95,70	69,74	8,94	76,60	12,54	56,33	92,23	71,44	85,33	24,60
6c: 1 + Indice Set 03	80,40	96,03	69,35	7,31	80,97	12,76	54,55	90,48	68,35	75,87	25,55
7: ”Good Choice”	81,64	96,97	70,33	8,10	84,73	11,39	58,33	91,70	66,54	85,73	29,63

Index set 1 including NDVI, BSI and NDWI resulted in a slightly lower overall accuracy in Vayots Dzor and a slightly higher overall accuracy in Sisian-Goris. Regarding the user’s accuracy of the target classes, fruit orchards, field crops and hay meadows increased in Vayots Dzor. In Sisian-Goris, Index set 1 enhanced user’s accuracy for improved meadows, field crops and fruit orchards, but slightly deteriorated user’s accuracy for hay meadows.

An inclusion of tasselled cap metrics as the second set of indices yielded in a slightly higher mean user’s accuracy of the target classes for both sub-regions. However, the improvement by the tasselled cap inclusion is quite low compared to the baseline. The combination of five indices in set three most notably enhanced the user’s accuracies of both field crops and fruit orchards as

well as the forest class for Vayots Dzor. Other classes showed only marginal effects. This is contrasting to the user's accuracy results in Sisian-Goris. Here the third set did yield slightly smaller mean accuracies for the target classes than the first two sets of indices. However, the mean accuracy estimates still improved in comparison to the baseline.

A beneficial combination of indices in terms of user's accuracy was determined through the analysis of the effect on the mapping reliability as a priority objective for thematic maps. This combination of indices yielded the highest mean user's accuracy for both sub-regions. The combination can be obtained from figure 10. Combined with the topography metrics and the nighttime lights, this set of indices was used for the "good choice" dataset for Vayots Dzor. Although the optical texture metrics did increase the user's accuracies of some target classes it was not considered in the "good choice" dataset because the effect was not given in the combined dataset as opposed to the simple addition to the baseline dataset. The same applies for the SAR time series in Vayots Dzor. Based on the differences between the sub-regions regarding the contribution of the different datasets, the "good choice" dataset was adjusted to Sisian-Goris. Compared to Vayots Dzor, nighttime lights were excluded and optical texture metrics added. Despite the deterioration of accuracies by the topographic data, the data was kept due to apparent visual improvements of the map prediction.

Compared to the baseline, the map reliability of the "good choice" classification in Vayots Dzor increased most for field crops and fruit orchards with about 6% as well as forest, urban and bare/ sparsely vegetated surfaces with almost 10%. Hay meadows, improved meadows and vine yards improved by about 3%, 4% and 1%, respectively. User's accuracies of grassland and shrubland deteriorated slightly by about 2%. In Sisian-Goris, the enhancement was most pronounced for improved meadows with 13%, and fruit orchards with 7%, sparse vegetation with 9% and urban fabric with 12%. The user accuracy decreased marginally for grassland and shrubland.

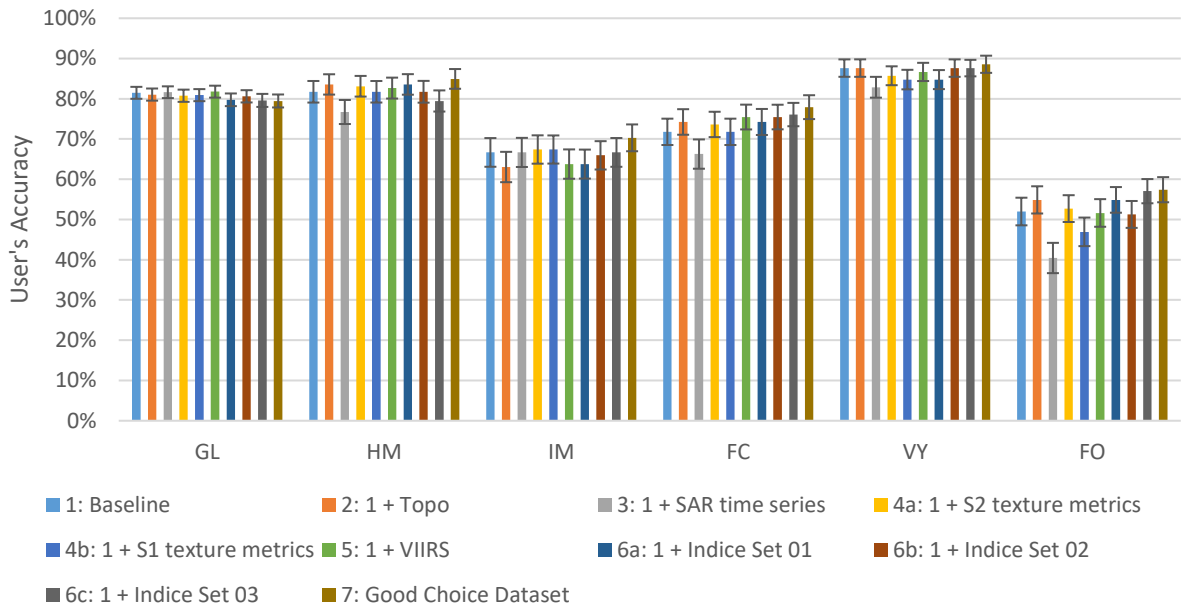


Figure 20 Bar chart illustrating the differences of the user's accuracies between the different RF models for Vayots Dzor. The standard error estimates are added as whiskers. A full version of this chart with all LULC can be obtained from appendix XXI.

Generally, the different dataset combinations show only small variations regarding the user's accuracy as priority metric (see figure 20 & 21). The improvements of the user's accuracy do not deliver a clear trend across the two sub-regions and show certain variations regarding the added value of the multisource data. Whereas in Vayots Dzor small improvements of different accuracy metrics could be observed for the ancillary topographic and nighttime lights data as well as specific sets of spectral derivates, the SAR time series and texture metrics lead to a reduced precision of the map predictions across all accuracy metrics. The Sentinel-2 texture metrics contribute to a small extent to the user's accuracy of the map. In Sisian-Goris however the optical texture and the SAR time series do improve the mean user's accuracy of the target classes. The topography metrics don't show an unambiguous effect within the accuracy and the nighttime lights deteriorate mapping accuracies. The spectral derivates such as NDVI or the Tasseled Cap transformation are functioning as indicators for different surface conditions and consolidate spectral information of the original bands. As a result, single indicators tend to enhance the accuracy of some classes only whereas combinations might result in synergistic effects. The precision of the user's accuracy estimates as provided by the standard errors is further limiting the relevance of the variations in the accuracy estimates between the different input datasets of the RF models (see figure 20 & 21).

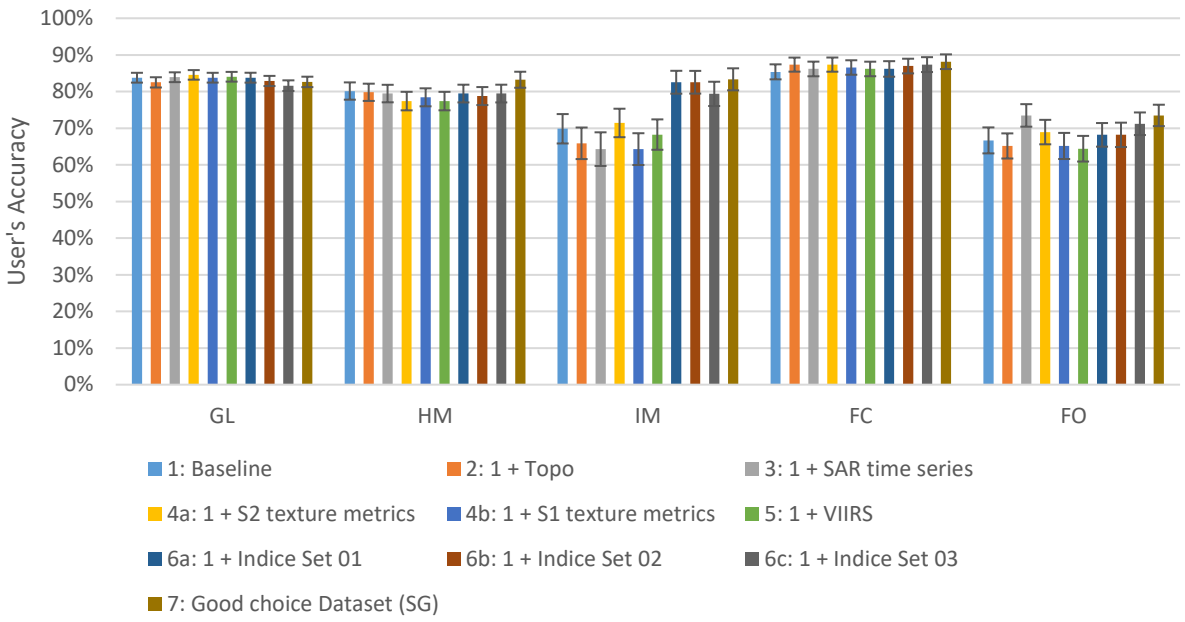


Figure 21 Bar chart illustrating the differences of the user's accuracies between the different RF models for Sisian-Goris. The standard error estimates are added as whiskers. A full version of this chart with all LULC can be obtained from appendix XXII.

The map of figure 22 shows small-scale representations of different areas (a-g) illustrating apparent misclassifications for both the baseline (top tile) and the "good choice" prediction (bottom tile). Despite of a good user's accuracy of grassland, the most abundant land cover class shows apparent spectral confusion with primarily shrubland but also with vine yards and the field crop class (see figure 22; a, b & g). The sparse vegetation/ bare class exhibits a continuum with the grassland and shows similar signs of confusion although more pronounced regarding the confusion with classes associated with a stronger spectral soil signal like vine yards, field crops and urban as opposed to shrubland. The inclusion of the topography metrics, the optical texture metrics as well as indices and nighttime lights mitigate these misclassifications to some extent. However, the spectral confusion between grassland and shrubland seems to be more persistent (see figure 22; b).

The topography metrics appear to effectively suppress the confusion in areas with the highest elevation and steepest terrain. The optical texture metrics, nighttime lights as well as the built-up index contribute to a reduction of confusion between urban and bare surfaces (see figure 22; a & e). Spectral confusion between shrubland and fruit orchards is further mitigated by vegetation indices and optical texture metrics (see figure 22; c). The reservoirs in the study region are prone for confusion with field crops due to temporal variability of the water level. This is also

partially mitigated by the inclusion of vegetation indices and topography metrics (see figure 22; c). The topography metrics as well as the indices also enhance the continuity of the field crop class from a more noisy prediction with gaps towards a more coherent classification result (see figure 22; f).

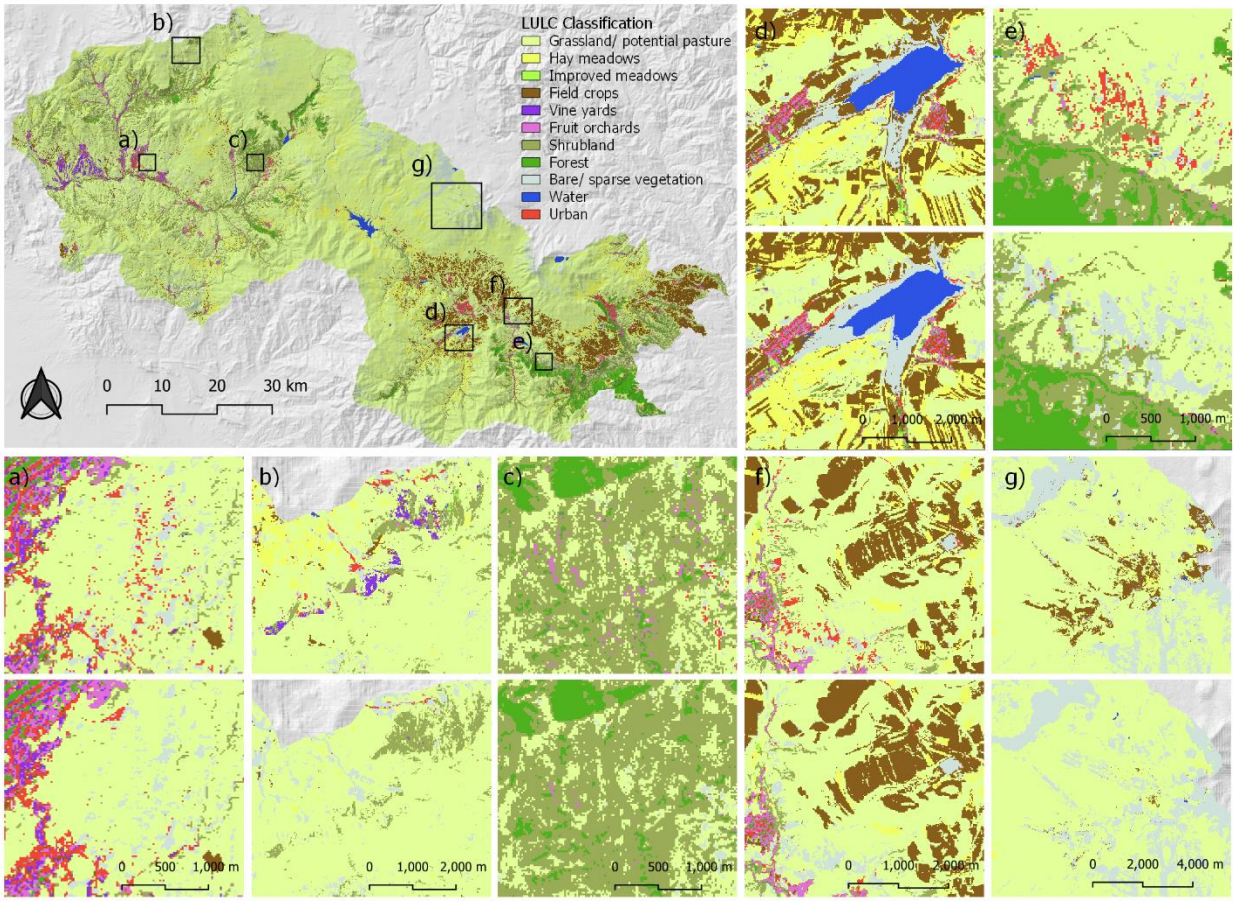


Figure 22 Map with small-scale representations of different areas (a-g) illustrating apparent misclassifications and enhancment for both the baseline (top tile) and the "good choice" prediction (bottom tile) (source: own illustration).

3.4.Spatial autocorrelation

Spatial autocorrelation could not be analysed for the whole study region due to computational constraints. Instead, seven ROIs of 30 km x 30 km were defined across the study region that cover different landscapes encountered (see appendix XI). Moran's I statistic was calculated for each ROI and plotted against pixel distance (figure 23). ROI 1 covers valleys with clusters of vine yards and fruit orchards, ROI 2 a landscape of grassland and hay meadows, ROI 3 shrubland and grassland, ROI 4 grassland, ROI 5 valleys with settlements, field crops and hay meadows, ROI 6 forest, shrubland and crop land, and ROI 7 large agricultural fields.

For some of the ROIs, values for Moran's I are generally higher and decrease less pronounced with increasing distance. This is in particular the case for the landscapes of steppe and alpine grasslands (ROI 4) as well as for shrubland (ROI 3) and, more surprisingly, also in the case of forest and cropland (ROI 6) and a heterogenous landscape with settlements, field crops and hay meadows (ROI 5). Contrastingly, other hay making areas (ROI 2) show lower values of Moran's I that decrease more pronounced with increasing distance. Against the initial expectation, the clustered areas of vine yards and fruit orchards (ROI 1) and the extended field crop landscape (ROI 7) also show low spatial autocorrelation.

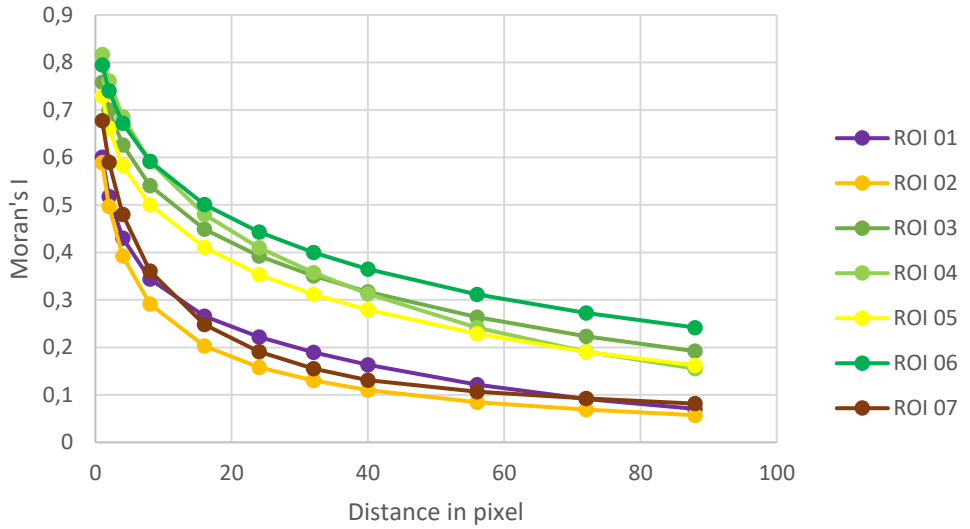


Figure 23 Moran's I values for the different environments (ROI 1-7) plotted against increasing pixel distance.

From the curve, Moran's I value ranges can be derived for a certain distance between pixels. For neighboring pixels with a distance of 0 pixels, the range of Moran's I is between 0.6 and 0.8 for all ROI. With increasing distance, the minimum and maximum values decrease, while the range increases first, and from a distance of about 20 pixels, decreases again.

The Moran's I range is used to derive the spatial autocorrelation present in the validation sample. In this study, the minimum distance between validation samples was set to 500 m for all classes, except for vine yards and improved meadows, where it was set to 250 m as their limited distribution otherwise wouldn't allow to sample enough points. As one pixel corresponds to 20 m distance, this results in a Moran's I range of 0.15 to 0.44. It can be expected, that for the more widely distributed grassland, shrubland or hay meadows, spatial autocorrelation with values around 0.4 is affecting the validation sample. In the case of more heterogeneously distributed

classes, spatial autocorrelation of around 0.15 is can be expected in the validation sample. For vine yard and improved meadow validation samples, Moran's I has values of around 0.2. The results show that the value ranges of spatial autocorrelation for the given validation data are within the limits of a low to moderate positive spatial autocorrelation and support a reliable accuracy assessment.

4. Discussion

The proposed classification system delivers an accurate LULC map with a detailed representation of the complex smallholder farming landscapes within the study region. Based on comprehensive classification trials the effects of multisource data were evaluated to derive a "good choice" input dataset with the most favourable map reliability across LULC classes. The optical bands alone did already achieve good accuracies and apparent confusion could be further mitigated using spectral derivates and ancillary data. The effect of the multisource data regarding the added value for classification are however relatively small. This indicates the limitations in respect to other factors such as the spatial or temporal resolution. The proposed simple classification system exhibits practicality regarding its application and some potential regarding its transferability. The produced map product provides information on the distribution and the extent of the LULC classes and indicates the nature value of the landscapes. The classified data further shows great relevance in respect to urging land-use challenges in the Armenian context. Besides these certain potentials of the applied mapping approach the results of this study indicate several limitations regarding the classification of mountainous and complex smallholder farming landscapes. Some of the limitations may be however overcome in a rather uncomplicated fashion by adapting the classification system without sacrificing the simplicity and the practicality.

In the following, the results of the LULC classification are discussed elaborating on potentials and limitations of the proposed mapping approach. The mapping accuracies of the classification trials and the added value of multisource data for map predictions are evaluated. The limitations regarding the spatial, temporal and spectral resolution are further addressed and put in context with significant mapping efforts of other studies. Implications related to the dimensionality of the RF input dataset and limitations regarding both training and validation data are noted and suggestions to overcome such limitations are provided. Finally, the practicality and transferability

of the classification system is assessed and some applications of the LULC product are reviewed in the context of current land-use challenges in Armenia.

4.1.Mapping accuracies

Accuracy of agricultural target classes

With an overall accuracy of 76% and 82% for Vayots Dzor and Sisian-Goris respectively the map predictions of the "good choice" RF model are far from random. However, the producer's accuracies and user's accuracies vary substantially across classes due to the spectral narrowness and high collinearity of the target classes. Despite mapping limitations indicated by low producer's accuracies, the proposed approach delivers user's accuracies ranging from 57% to 88% in both regions. The low accuracies of some classes need to be put in context with regular mapping constraints of difficult classes such as transitional classes (e.g. shrubland). The mapping limitations causing low accuracies are discussed for specific classes in the following.

Sparse woody vegetation is usually associated with a pronounced herbaceous layer. The collinearity of transitional classes with both a signal of woody vegetation and an herbaceous signal seems to limit a more accurate distinction. Irrigation contributes to a clearer separability of managed classes such as fruit orchards and improved meadows from unmanaged classes, but hampers a more accurate distinction between irrigated classes due to the strong herbaceous reflectance signal. Varying successional stages of the woody vegetation within classes further contributes to this. This can be observed for fruit orchards where more recent plantings are hard to discriminate from improved meadows.

Due to the extensive management, spectral signals of smallholder agriculture are usually weak resulting in a poor separability across management regimes. Nevertheless, hay meadows and field crops show lower errors of commission and omission of around 10-30%. Yet, there is a very clear difference of the producer's accuracy between the two regions for field crops with 18.5% in Vayots Dzor and 84.7% in Sisian-Goris. This could be explained by more intensive cultivation management in parts of Syunik with larger farm sizes and larger fields as opposed to marginal and extensive cropping of grains and vegetables in Vayots Dzor (see figure 6 in section 2.1). The increased class dimensionality in Vayots Dzor in combination with these differences in

farming systems is most likely contributing to the differences in overall but also user's and producer's accuracies between both regions.

Added value of multisource data

Although still prevalent in the "best choice" classification, the inclusion of multisource data could partially address these limitations and improved user's accuracies of some target classes.

Regarding the use of spectral derivates, the results of this study appear to be in line with the literature. As in a case study of orchard mapping in the Maipo valley in Chile, the accuracy estimates for the different datasets suggest that a small number of indices in addition to the original optical bands do not improve the accuracy as opposed to using the optical spectral bands only (Peña, Liao and Brenning, 2017). In fact, the baseline already yields quite satisfactory accuracies. Yet, when using a combination of many indices in addition to the original spectra, accuracies improve more substantially. Additional to the explorations presented in this study, trials could be performed using only indices without optical spectra or trials that include additional band spectra of the Sentinel-2 sensor (i.e. the red edge bands).

Against initial expectations, the combination of optical and SAR imagery did not improve mapping accuracies. Both regions show small variations in respect to this effect. A slight deterioration in Vayots Dzor and a marginal improvement in Sisian-Goris can be observed. The accuracies of both regions suggest that the SAR time series may be beneficial for the mapping of sparse woody vegetation as it consistently enhanced user's and producer's accuracies for shrubland in both regions. This might also explain the increased accuracy of fruit orchards in Vayots Dzor, where scattered fruit and nut trees are a dominant land use. This would also verify suggestions from former studies regarding the added value of a combination of optical and SAR data (Van Tricht *et al.*, 2018; Brinkhoff, Vardanega and Robson, 2020). Possible factors which might limit a beneficial effect of the complementary information provided by combined data could be either the high dimensionality of the input dataset or the low power of discrimination due to insufficient pre-processing of the SAR data.

The use of ancillary data boosted classification accuracies depending on the data type and the specific class. The topography metrics show a high RF feature importance and descriptive statistics suggest that the inclusion is beneficial for the distinction of managed and semi-natural

classes. However, mapping accuracies do not show clear tendencies with only marginal improvements in Vayots Dzor and slight deterioration in Sisian-Goris. Despite these findings, the use of topography metrics in complex terrain can be encouraged to complement optical imagery in RF models. This is based on previous work (e.g. Adepoju and Adelabu, 2020) and apparent visual effects within this study. It is possible that the discriminative power of the topography metrics is lower, with a larger training sample and highly dimensional input data resulting in a reduced explanatory importance of the topography metrics. The nighttime lights showed a beneficial effect regarding the distinction of urban structures from bare or sparsely vegetated land cover and fallow fields. In case of the texture metrics, only the optical texture metrics had a positive effect on the map reliability. The SAR texture could be affected by backscattering or compositing issues. The effect of texture might generally be limited due to insufficient spatial resolution (10-20m) to characterize important structures, of perennial crops for example. It should be tested in future studies, whether texture metrics derived from VHSR imagery could overcome such limitations.

Overall, the combination of optical spectra and derivates as well as radar and other multisource data only had limited effects on the performance of the RF predictions. Certain classes benefited whereas other classes faced drawbacks. When combining different beneficial input data within a “good choice” dataset, synergies can be used to attain more profound improvements of model predictions with special regards to the user’s accuracy as priority mapping objective. In this case, the improvements of user’s accuracies between the baseline dataset and the “good choice” dataset have a limited range from -1% to +13% with ±3% precision. These limitations together with partially low producer’s accuracy reflect the difficulties of mapping smallholder land use in a complex mountainous landscape. Due to the small variations across the datasets, the differences were not tested for statistical significance in this study which may be a certain shortcoming. Although the precision of the accuracy estimates can indicate if the difference is significant, there is no statistical proof.

4.2.Limitations and potential of classification system

The low effect of the multisource data inclusion highlights other possible limitations regarding the classification approach and RF input data. This mainly includes considerations of

increased spatial, temporal and spectral resolution, but also whether to use a pixel- or object-based approach, to apply a more sophisticated classifier, to use a hierarchical and more automated mapping approach or to improve the collection of training data. In the following this will be discussed briefly and put in context with the results of this work and other similar studies. The contextual view to other studies can however not be interpreted as comparison given the multitude of methodologically varying LULC classification approaches. The practicality and transferability of the applied system are further evaluated. Table 8 summarizes advantages and disadvantages of the proposed mapping approach.

Table 8 Advantages and disadvantages of the proposed mapping approach.

Advantages	*Disadvantages*
Open source data & software	Complicated validation sampling scheme
Cohesive LULC mapping (target & by-product classes)	Low degree of automation & labour intensive
Information on class specific confusion	Interpreter bias & lack of ground truth
Relatively simple and effective classification system	Accuracy limitations
No need for detailed ground truth assessment	Limited typological detail
Benefits of multisource and ancillary data	Unexplored spectral, temporal and spatial potential of Sentinel-2
Transferability & adjustability	High dimensionality of dataset

Considerations of spatial resolution

The spatial resolution seems to be a major limiting factor for an accurate and detailed map representation of mountainous smallholder farming in Armenia. The highly fragmented and patchy land use, small field sizes and mixed farming activities are clearly putting constraints on the classification in some areas. This is especially the case for the home gardens and small fields in settlements and valleys of the study region.

The minimum mapping unit within this study is one pixel, following the spatial resolution of the optical Sentinel-2 spectra (20m). The smallest feature that could be possibly mapped would be therefore 20 m x 20 m or 0.04 ha. A feature with a size of 0.01 to 0.05 ha may not fall entirely

within one given pixel and could be split among as many as four pixels. The feature would therefore only make up a minority of those pixels with no sufficient dominance to be sensed clearly. In case of smallholder farming with field sizes as small as 0.05 hectares, the minimum mapping unit would need to be 10 m x 10 m pixels to allow a 4x4 pixel unit to entirely contain such small features (Herold, 2011).

In respect to spatial resolution Sentinel-2, offers unexplored potential providing the visual RGB spectra and a broader NIR band with a spatial resolution of 10 meters. These bands may be used to pan-sharpen the other spectral bands (Verma and Jana, 2019). With many simple NDIs using the red and infrared bands only, the proposed mapping approach could be easily adapted to this increased resolution. Yet, the lack of multiple spectra might also limit mapping improvements. Other VHSR sensors mainly acquire images in the RGB and NIR spectra, but the increased spatial and temporal resolution has high potential for mapping efforts. In case of VHSR imagery, object based approaches with some sort of spatial and semantic segmentation are often necessary to take full advantage of their spatial and contextual information (Lebourgeois *et al.*, 2017; Du *et al.*, 2019).

Considerations of temporal resolution

The monthly temporal resolution of the time series used in this study may be too coarse for a detailed representation of the phenological differences and variations in management regimes between classes. The detection of the mowing events for hay meadows can be hampered through the low temporal resolution possibly blurring the management signal. Depending on cutting intensity and recovery rate of the vegetation this issue could be amplified (Griffiths *et al.*, 2020). Vice versa, a reduction of biomass due to livestock grazing for example could be mistaken with the abrupt cutting intervention with insufficient temporal detail. This is in line with findings from studies on grassland use and use intensities regarding hay making in Switzerland (Kolecka *et al.*, 2018) and Germany (Griffiths *et al.*, 2020). Kolecka *et al.* (2018) mapped 4 different grassland use intensities at a parcel level using Sentinel-2 data and correctly detected 77% of the mowing events. They showed that a lower temporal resolution diminishes accuracies and that parcels with high mowing frequencies are especially sensitive to a removal of image dates. In Germany, Griffiths *et al.* (2020) mapped 5 different mowing intensity classes using the harmonized Landsat-Sentinel dataset and observed higher omission errors for pixels that only experienced a single

mowing event. They emphasize the importance of temporal resolution and short compositing intervals for the time series to not omit mowing events.

In respect to the five days revisit time of Sentinel-2 and the relatively low cloud contamination, the optical time series in this study could be calculated with an increased temporal density (e.g. 10-day composites) to cope with this limitation. Many recent detailed LULC studies use short temporal binning intervals and dense time series. Griffiths, Nendel and Hostert (2019) used the harmonized Landsat-Sentinel dataset from NASA and derived a gap-filled dense 10-day time series to map 12 different agricultural and semi-natural classes across Germany and achieved an overall mapping accuracy of 81%. Rufin *et al.* (2019) explored time series binning for cropping practice discrimination in Turkey, showing that quarterly binning of spectral-temporal metrics yields higher mapping accuracies compared to annual binning. They also produced a dense gap-filled 8-day time series of the Tasseled Cap components. Both approaches resulted in overall accuracies exceeding 90%.

Considerations of spectral resolution

Finally, there is some unexploited potential regarding the spectral resolution within the proposed classification system. In this study, only the narrow NIR band (band 8A) was considered for the calculation of the spectral derivates and as band input for the different layerstacks. However, Sentinel-2 offers an additional broader NIR band and red edge bands. This multi-spectral ability in the near infrared could be used to further enhance the mapping accuracy. Griffiths, Nendel and Hostert (2019) showed that an inclusion of the red edge band of Sentinel-2 and a simple proxy for the missing red edge observations of Landsat-8 improves mapping performance for both semi-natural and crop classes.

Limitations of dimensionality

With original spectral bands, spectral derivates, derived spectral-temporal metrics and texture metrics as well as the monthly time series and ancillary data, the dimensionality of the input dataset is likely to limit the ability of the RF model to accurately predict the LULC classes. Although RF is largely resistant to overfitting and generally capable to process large amounts of multivariate or uncorrelated data, an increase in the dimensionality of the input dataset may require a substantially larger training sample size in order to ensure sufficient accuracies (Belgiu

and Drăgu, 2016). Feature selection on the basis of feature importance or a Principle Component Analysis (PCA) are widely used to reduce data dimensionality and overcome negative effects on mapping accuracies. The provided feature importance of the RF models in GEE is a mean to inform about dimensionality issues (see appendix XVIII), however, a simple reduction of the input dataset based on this feature importance alone was avoided because it was not clear how it is derived. The calculation of a PCA, a spectral rotation that takes spectrally correlated image data and outputs uncorrelated data, was limited by the size of the datasets within GEE and therefore not applied (GEE, no date a).

Limitations of typological detail, training and validation data

The field work restrictions due to the COVID-19 pandemic caused certain limitations regarding the thematic detail of the class typology and uncertainties for both training and validation data. The lack of reference data is limiting both increased typological detail and unbiased accuracy and area estimates. Because the methodology is designed to acquire training and validation data by simply interpreting remotely sensed data, the approach may face serious challenges in respect to uncertainties of class membership assignment. This largely depends on the experience and skills of the interpreter and level of detail required by the classification system (Congalton and Green, 2019). Interpreter bias is therefore expected to have some impact on the mapping efforts of this study.

The proposed classification system was primarily designed to cope with these limitations and measures have been implemented that ensure coherent and valid training and validation samples. Without the immediate necessity of field work, the assessment of class membership is solely based on remote assessment and visual interpretation of multi-temporal VHSR images as well as other available spatial-temporal data. The independence from detailed reference data or ground visits can also be considered a strength of the proposed class typology. The thematic detail could be further increased with data gathered in field trips. Given the importance of the training and validation stage to direct the classification and evaluate the map prediction, the collection of training and validation data was applied carefully with considerations regarding a right balance, class specific representativeness and spatial autocorrelation to ensure a comprehensive collection scheme and high-quality samples.

The collection of training data within this study is very time consuming. RF generally requires a large training sample for highly dimensional datasets. The work-load may be reduced by implementing a sequential mapping approach using thresholds and image masking. However, the development of such approaches usually requires a-priori available geospatial and spectral data to explore thresholds and conceptualize a sequential masking procedure following certain assumptions. This could be applied in future work, possibly building upon the gathered and produced data of this study.

Trade-off: practicality and ease of application vs. mapping accuracy

The proposed simple pixel-based classification system makes use of pre-processed and readily usable intra-annual multi-temporal and multi-spectral optical and SAR data as well as ancillary data within GEE. This offers great practicality which can be considered advantageous compared to the use of VHSR imagery as well as complex and labour-intensive object-based or hierarchical approaches. The trade-off between workload and accuracy is becoming clearer though when comparing the rather simple classification system of this study to more sophisticated approaches in other studies. Lebourgeois *et al.* (2017) for example applied a hierarchical, object-based mapping approach of smallholder agriculture with simulated Sentinel-2 data in Madagascar. Despite the high degree of typological detail, they achieved high accuracies across crop and non-crop domains. With increasing detail, the accuracy only decreased substantially for the crop domain, highlighting the difficulties of mapping smallholder agriculture. Van Tricht *et al.* (2018) also applied a hierarchical classification approach combining optical and radar data to map 12 crop classes and three by-product classes in a complex agricultural landscape in Belgium, achieving a maximum overall mapping accuracy of 82%.

Object-based classification is commonly considered to provide improved performance compared to pixel-based classification (Stefanski *et al.*, 2014; Brinkhoff, Vardanega and Robson, 2020). However, spatial segmentation is time consuming and the application for agricultural mapping is mainly used for intensively managed landscapes with larger field sizes. Pixel-based approaches are preferred for fields that are around or smaller than a pixel (Jain *et al.*, 2013). In regions where small-scale agriculture is prevalent, pixel-based methods may be advantageous or equal in terms of accuracies compared to object-based analysis (Yin *et al.*, 2018, 2019). Rufin *et*

al. (2019) emphasizes that a pixel-based approach is not restricted by MMUs and allows capturing small parcels in fragmented landscapes.

The necessity of spatial segmentation in classification systems based on VHSR imagery is limiting the applicability. Another limitation for the practicality is that VHSR imagery is mostly only free of charge with a low degree of pre-processing. This requires additional workload as opposed to the use of readily usable data products provided in GEE. In order to handle the large data amount and the spatial complexity of VHSR imagery, more sophisticated classification approaches are often applied. Besides deep learning applications, model stacking can be used to assess the LULC of smallholder-dominated and complex landscapes in VHSR imagery (Aguilar *et al.*, 2018; Du *et al.*, 2019). Aguilar *et al.* (2018) mapped five different crops in a smallholder farming landscape in Mali achieving an overall accuracy of 76% from a prediction of 75 classifiers mining multi-temporal optical data from Worldview-2.

The proposed classification approach delivers an all-inclusive representation of the mountainous smallholder farming landscape. The cohesive character of the class typology provides spatially explicit representations of both target classes and by-product classes without the need of an empty or undefined class (i.e. “other” LULC). This requires the allocation of all pixels to the specified classes, even if only marginally separable. The consideration of coherent classes in order to provide “the entire picture” increases the class dimensionality and thus diminishes the mapping accuracy. Such cohesive and holistic mapping efforts can be however beneficial for applied research.

Besides supervised classification, LULC datasets can be created by manual digitizing as in the case of the CORINE land cover. It is de facto the standard for LULC monitoring at the pan-European level and the national teams of experts are assessing the LULC based on the CORINE nomenclature (44 classes), an MMU of 25 ha and varying methods (Aune-Lundberg and Strand, 2021). There is a recent pilot project to extend the CORINE database in the eastern partnership countries including the Caucasus (EEA, 2016). Although the dataset exhibits a mapping accuracy of around 87%, the low resolution usually makes it less relevant for use at the national level. The nomenclature further insufficiently distinguishes between land cover and land use and the MMU makes it necessary to use mixed classes and omit detail (Aune-Lundberg and Strand, 2021).

Considering the given context of detailed LULC mapping the limitations and potentials of this study become clearer. Many of the above-mentioned studies achieved higher accuracies due to more complex but thus also more labour-intensive approaches such as image fusion, model stacking, hierarchical approaches, dense gap-filled time series, object-based methods and the use of VHSR imagery. Yet, some practical measures which could be considered to improve the classification system without sacrificing its potential regarding the applicability and simplicity have also been identified. This includes quarterly binning of spectral-temporal metrics and a 10-day time series, the use of red edge spectra as well as optical bands with a spatial resolution of 10 meters and feature selection for RF.

Transferability of the classification system

The classification system exhibits some potential regarding the transferability to other smallholder dominated and complex landscapes as well as more intensively used agricultural areas. As demonstrated for the two sub-regions of this study, both the class typology and the dataset can be adjusted to the specific situation. Although the mapping accuracy is decreasing with increasing class dimensionality, the flexibility of the typology allows the aggregation of classes in case the thematic detail was too high to achieve satisfactory accuracies. Thus, on a municipal level, the approach can be transferred and adjusted with a minimum detail level as baseline and a highly thematic ceiling. The class typology is largely generic and follows basic assumptions about management systems as well as commonly used LULC classes, supporting transferability.

The cropland and hay meadow classes for example are based on the spectral-temporal signal of the management interventions of ploughing and hay making. This management is commonly practiced and the associated signal may be sensed in any region. Further, classes as hay meadows can also be condensed by aggregating all mowing intensities into one class, or thematically increased by differentiating different mowing frequencies as different classes and hence linked to classification systems on broader scales (Verma and Jana, 2019). For the area of this study, a differentiation between the extensively used hay meadows (cut one time) and the intensively used and irrigated improved meadows seemed sufficient.

The stratification of the study region in agroecological zones was done in order to ensure a comprehensive training sample and achieve high mapping accuracies. Stratification may put

constraints on transferring the approach to smaller geographical scales (e.g. continental), but could be beneficial for regional studies at a larger scale. It can be in particular interesting for areas similar to the study region featuring mountainous and heterogenous landscapes with a high spatial and spectral complexity. This can be useful to overcome classification limitations in respect to spectral and spatial differences along environmental gradients and guarantee accurate map predictions across regions. With a stratification based on municipal districts, the data can be analysed at a greater depth enabling customized options for local decision makers. The main objective within this study was a high accuracy and detail for a map representation on a regional scale. This could be ensured by the minor adjustments of the classification system in terms of RF input data and class depth. Yet, such adjustments contribute to a less intuitive and more labour-intensive transferability of the mapping approach.

The application of the exact same classification system across different regions can be challenging regarding variations of spatial complexity and may require an adjustment of the system (Hast and Mehari, 2016). Training-signature-based classification systems generally face limitations regarding transferability as they require repetitive training sampling for every region or scene. This is very tedious and time-consuming and a disadvantage compared to rule-based classification processes. However, depending on the complexity of the data (i.e. VHRS) and sensed environments, defining a fully reliable rule set may be significantly limited due to spectral and spatial differences (Hussain and Shan, 2016). For less complex data, rule inheritance of classification systems and learning transferability (i.e. use of labeled data from auxiliary domains for learning in another domain where labeled data is scarce or absent) can be promising ways forward to provide a better model transferability (Hussain and Shan, 2016; Sukhija, Krishnan and Kumar, 2018).

4.3.Applications of LULC product

As described previously, detailed mapping of smallholder agriculture in Armenia has a special relevance to address current challenges. This particularly concerns the livelihoods of the rural population on the one hand and ecosystem services as provided by intact biodiversity of resilient mountainous ecosystem on the other hand. The produced LULC map delivers inherent information on the intensity within the agricultural target land use as well as class coherence with

a rich semi-natural land-cover by-product. This dataset may be further used to assess the nature value of mountainous farmland in Armenia. A simple application could be the assessment of spatial patterns of hay making areas such as comparing hay cutting close to settlements with more remote subalpine zones (see figure 18). More complex analysis, utilizing the full information level of the dataset, could be multi-criteria weighted overlays for spatial-explicit modelling of the nature value of smallholder dominated landscapes. The nature value of farmland may be a relevant concept to combine and balance the interests of agricultural production and landscape conservation.

Current plans regarding land use in Armenia require spatially explicit and detailed information at hand for policy- and decision-makers in order to carefully plan actions and evaluate potentials and trade-offs. Afforestation, land consolidation and sustainable intensification as well as wildlife conservation are urging tasks but decisions should be based on a comprehensive knowledge base of LULC and its dynamics. As part of the GAtES project, this study adds to previous project-studies on LULC mapping, energy use of rural communities, afforestation scenarios, and public participation mapping of ecosystem services (Harutyunyan, Petrosyan, *et al.*, 2019; Harutyunyan, Pfeiffer, *et al.*, 2019; Harutyunyan, Schlaffer, *et al.*, 2019; Schulte and Harutyunyan, 2020). The strong relation to land system science enables to connect to remote sensing work on post-soviet land-use dynamics of a research joint including scientists from the University of Copenhagen, the Humboldt University and the University of Wisconsin-Madison. This work may complement existing remote sensing work in the Caucasus on long term LULC change, habitat suitability of mountain ungulates or landscape connectivity (Bleyhl *et al.*, 2017; Buchner *et al.*, 2020; Kuemmerle *et al.*, 2020).

Wildlife management and corridor planning is of considerable importance within the study area and the high level of typological detail of the LULC product exhibits potential for corridor management and monitoring of different human activities affecting wildlife (see figure 19). For protected areas, land-use pressures could be assessed to direct decision making within planning processes for new protected areas. An example for a promising conservation project within the study area is the Arpa Protected Landscape, a community lead protected area which includes the communities of Khachik, Gnishik and Areni. The initiative uses contractual nature conservation agreements and a clear set of processes to fund ecologically sustainable land use. Thus, it seeks

to address land-use issues by supporting and empowering local communities and directly including them in conservation efforts (Kuntz, 2019).

Finally, the LULC data may contribute to improve livelihoods of the smallholder farmers and the efficiency of agricultural land use in Armenia. Land consolidation and sustainable intensification schemes can benefit from detailed LULC data. Since the classification system is applied with a pixel-based approach and not on plot level, the data cannot inform on fragmentation of the land use. However, the information on nature value of the mountainous small-scale farming systems can deliver valuable insights on a landscape level and may thus facilitate and direct decision making. It can be used as an indication of areas where sustainable intensification or land consolidation may make sense and where a conservation of the system due to a higher nature value might be preferred.

5. Conclusion

The proposed approach is a first step towards a detailed, spatially explicit assessment of LULC in the smallholder-dominated mountainous and complex landscapes of the Armenian Highlands. The LULC product delivers an accurate representation of the landscape's complexity and has great potential for various current challenges in Armenia regarding land use, biodiversity and ecosystem services. The classified map successfully indicates the nature value of the smallholder farming systems due to the detailed and cohesive character of LULC class typology featuring inherent information on land-use intensities and rich semi-natural and artificial by-product classes. The outcomes of LULC classification indicate the major constraints of assessing narrow and highly colinear classes of such a detailed class typology.

The mapping results further illustrate the degree of usefulness of the multisource data products to aid the LULC classification and improve accuracies. A combination of optical and SAR data didn't yield a general accuracy enhancement apart from a positive trend for woody vegetation. Ancillary data such as topography metrics, nighttime lights and optical texture metrics enhanced the map predictions partially and were therefore considered within the "good choice" classification. Single spectral derivates (i.e. NDIs) only favoured single classes while the inclusion of a set of derivates led to a more profound enhancement of mapping accuracies across classes. The overall effect of the multisource data on the mapping accuracies is however relatively low

and varies between the sub-regions. These limitations indicate that other factors such as spatial or temporal resolution may constrain higher mapping accuracies. An increased spatial and temporal resolution is likely to further improve the ability of RF to capture the small agricultural plots and the phenological differences between classes. This can be achieved in a rather uncomplicated fashion by constructing shorter temporal bins or using the Sentinel-2 bands with a spatial resolution of 10 m.

In general, the classic pixel-based classification approach is simple, practical and effective while still delivering promising results. In comparison with other, more complex classification systems, the proposed approach is straightforward and relatively easy to apply. It doesn't require detailed reference data, satellite imagery purchase, high computational power or expensive software. The system can be transferred and adjusted to other administrative districts of Armenia to provide full coverage. The limitations of the typological detail of the classes and interpreter bias related to the collection of training and validation data were largely overcome trough a comprehensive sampling scheme. This could be further improved with additional data or field assessments.

The LULC data produced in this study is rather technocratic and the knowledge generation took place distant to the local communities which potentially contributes to a feeling of exclusion. Therefore, it is of importance to combine these results with studies investigating the needs of smallholders and rural communities as well as the values they associate and perceive in respect to ecosystem services. Community led projects in Armenia such as the "Arpa Protected Landscape" can be a mean to empower local smallholders to actively contribute to conservation management and monitoring. Detailed LULC data is likely to be a valuable tool for such projects. Geodata processing and remote sensing applications are becoming more accessible for a broader spectrum of users due to cloud computing, open access data and available freeware. This also bears some potential for local authorities and specialists to further develop and apply the proposed mapping approach of this study.

References

Adepoju, K. A. and Adelabu, S. A. (2020) 'Improving accuracy evaluation of Landsat-8 OLI using image composite and multisource data with Google Earth Engine', *Remote Sensing Letters*, 11(2), pp. 107–116. doi: 10.1080/2150704X.2019.1690792.

Aghababyan, K. *et al.* (2016) 'First National Atlas of the Birds of Armenia', *Bird Census News: Journal of the European Bird Census Council*, 28(2), pp. 52–58. ISSN 1381-5261.

Aguilar, R. *et al.* (2018) 'A cloud-based multi-temporal ensemble classifier to map smallholder farming systems', *Remote Sensing*, 10(5). doi: 10.3390/rs10050729.

Alcantara, C. *et al.* (2013) 'Mapping the extent of abandoned farmland in Central and Eastern Europe using MODIS time series satellite data', *Environmental Research Letters*, 8(3). doi: 10.1088/1748-9326/8/3/035035.

Aleksanyan, A. *et al.* (2020) *Biodiversity of dry grasslands in Armenia: First results from the 13th EDGG Field Workshop in Armenia*. doi: 10.21570/EDGG.PG.46.12.

Amatulli, G. *et al.* (2018) 'Data Descriptor: A suite of global, cross-scale topographic variables for environmental and biodiversity modeling', *Scientific Data*, 5, pp. 1–15. doi: 10.1038/sdata.2018.40.

Artavan forest (no date) *Google Earth Photos*. Available at: https://lh5.googleusercontent.com/p/AF1QipP_oy8U0iupNNbrsHxVmipCqhmJLur7RZDgU74e=h1440 (Accessed: 10 August 2020).

Asatryan, A. (2017) *Introduction into the botanical paradise of Armenia*, *Wordpress*. Available at: https://gbportalnet.wordpress.com/2017/02/12/introduction-into-the-botanical-paradise-of-armenia/ (Accessed: 15 August 2020).

Asatryan, A. (2019) 'Wild Pears of Armenia', *Acta Biologica Plantarum Agriensis*, 7, pp. 19–31. doi: 10.21406/abpa.2019.7.19.

Askerov, E. *et al.* (2019) 'Leopard (Panthera pardus) reoccupying its historic range in the South Caucasus: a first evidence (Mammalia: Felidae)', *Zoology in the Middle East*, 65(1), pp. 88–90. doi: 10.1080/09397140.2018.1552349.

Aune-Lundberg, L. and Strand, G.-H. (2021) 'The content and accuracy of the CORINE Land Cover dataset for Norway', *International Journal of Applied Earth Observation and Geoinformation*, 96, p. 1-10. doi: 10.1016/j.jag.2020.102266.

Avetisyan, S. (2010) 'Agriculture and Food Processing in Armenia'. Limush Publishing House, Yerevan. Available at: tinyurl.com/3p7kqyca.

Batsatsashvili, K. *et al.* (2016) 'Thymus caucasicus Willd. ex Benth. Thymus collinus M. Bieb. Thymus kotschyanus Boiss. & Hohen. Lamiaceae', In: Bussmann R. (eds) Ethnobotany of the Caucasus. European Ethnobotany. Springer, Cham. doi: 10.1007/978-3-319-50009-6_101-1.

Baumann, M. *et al.* (2018) 'Mapping continuous fields of tree and shrub cover across the Gran Chaco using Landsat 8 and Sentinel-1 data', *Remote Sensing of Environment*, 216, pp. 201–211. doi: 10.1016/j.rse.2018.06.044.

Belgiu, M. and Drăgu, L. (2016) 'Random forest in remote sensing: A review of applications and future directions', *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, pp. 24–31. doi: 10.1016/j.isprsjprs.2016.01.011.

Berhane, T. M. *et al.* (2018) 'Comparing pixel- and object-based approaches in effectively classifying wetland-dominated landscapes', *Remote Sensing*, 10(1). doi: 10.3390/rs10010046.

Bleyhl, B. *et al.* (2017) 'Assessing landscape connectivity for large mammals in the Caucasus using Landsat 8 seasonal image composites', *Remote Sensing of Environment*, 193, pp. 193–203. doi: 10.1016/j.rse.2017.03.001.

Brinkhoff, J., Vardanega, J. and Robson, A. J. (2020) 'Land cover classification of nine perennial crops using sentinel-1 and -2 data', *Remote Sensing*, 12(1), pp. 1–26. doi: 10.3390/rs12010096.

Buchner, J. *et al.* (2020) 'Land-cover change in the Caucasus Mountains since 1987 based on the topographic correction of multi-temporal Landsat composites', *Remote Sensing of Environment*, 248. doi: 10.1016/j.rse.2020.111967.

CBD (2014) *Fifth National Report of Armenia to the Convention on Biological Diversity*. Available at: https://www.cbd.int/doc/world/am/am-nr-05-en.pdf.

Chen, Y. *et al.* (2018) 'Detecting irrigation extent, frequency, and timing in a heterogeneous arid agricultural region using MODIS time series, Landsat imagery, and ancillary data', *Remote Sensing of Environment*, 204, pp. 197–211. doi: 10.1016/j.rse.2017.10.030.

CLIMATE-DATA.ORG (no date) *Armenia Climate*. Available at: https://en.climate-data.org/asia/armenia-10/ (Accessed: 2 October 2020).

Cochran, W. G. (1977) 'Sampling techniques third edition'. Wiley Series in Probability and statistics. ISBN: 978-0-471-16240-7.

Collison, A. (2017) *A Clear Look at the Earth's Surface: Planet's Surface Reflectance Beta*. Available at: https://www.planet.com/pulse/a-clear-look-at-the-earths-surface-planets-surface-reflectance-beta/ (Accessed: 12 November 2020).

Congalton, R. G. and Green, K. (2019) *Assessing the Accuracy of Remotely Sensed Data*, *Principles and Practices*. Third Edition. CRC Press. doi: 10.1201/9780429052729.

Curtis, G. and Library of Congress (1995) *Armenia, Azerbaijan, and Georgia: country studies*. Federal Research Division, Library of Congress. Washington, D.C. ISBN 0-8444-0848-4.

Das, T. and Das, A. K. (2014) 'Mapping and Identification of Homegardens as a Component of the Trees Outside Forests Using Remote Sensing and Geographic Information System', *Journal of the Indian Society of Remote Sensing*, 42(1), pp. 233–242. doi: 10.1007/s12524-013-0310-3.

Dash, J. and Ogutu, B. O. (2015) 'Recent advances in space-borne optical remote sensing systems for monitoring global terrestrial ecosystems', *Progress in Physical Geography*, 40(2), pp. 322–351. doi: 10.1177/0309133316639403.

Davtyan, A. (no date) *Syunik*, *Google Earth Photos*. (Accessed: 10 August 2020).

Diek, S. *et al.* (2017) 'Barest Pixel Composite for agricultural areas using landsat time series', *Remote Sensing*, 9(12). doi: 10.3390/rs9121245.

Du, Z. *et al.* (2019) 'Smallholder crop area mapped with a semantic segmentation deep learning method', *Remote Sensing*, 11(7). doi: 10.3390/RS11070888.

Dusseux, P. *et al.* (2014) 'Agricultural practices in grasslands detected by spatial remote sensing', *Environmental Monitoring and Assessment*, 186(12), pp. 8249–8265. doi: 10.1007/s10661-014-4001-5.

EEA (2016) *Implementation of the Shared Enviromental Information System (SEIS) principles and practices in the ENP East region (ENI SEIS II East)*. Available at: https://eni-seis.eionet.europa.eu/east/areas-of-work/data/land (Accessed: 10 January 2021).

Elbakidze, M. and Angelstam, P. (2011) 'Implementing sustainable forest management in Ukraine's Carpathian Mountains: The role of traditional village systems', *Forest Ecology and Management*, 249(1), pp. 28-38. doi: 10.1016/j.foreco.2007.04.003.

ESA (2015) *Sentinel-2 User Handbook*. Available at: https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook (Accessed: 11 November 2020).

ESA (no date a) *Level-2A Algorithm - Sentinel-2 MSI Technical Guide - Sentinel Online*. Available at: https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm (Accessed: 11 November 2020).

ESA (no date b) *Sentinel-1 - Data Products - Sentinel Online*. Available at: https://sentinel.esa.int/web/sentinel/missions/sentinel-1/data-products (Accessed: 12 November 2020).

ESA (no date c) *Sentinel-1 - Instrument Payload - Sentinel Online*. Available at: https://sentinel.esa.int/web/sentinel/missions/sentinel-1/instrument-payload (Accessed: 12 November 2020).

ESA (no date d) *Sentinel-1 - Missions - Sentinel Online*. Available at: https://sentinel.esa.int/web/sentinel/missions/sentinel-1 (Accessed: 12 November 2020).

Estel, S. *et al.* (2018) 'Combining satellite data and agricultural statistics to map grassland management intensity in Europe', *Environmental Research Letters*, 13(7). doi: 10.1088/1748-9326/aacc7a.

FAO (2016) 'Map accuracy assessment and area estimation : a practical guide', *National forest monitoring assessment working paper*, E(46), p. 69. Available at: http://www.fao.org/3/a-i5601e.pdf.

FAO (2020a) *FAO addresses agricultural land abandonment in Armenia*. Available at: http://www.fao.org/europe/news/detail-news/en/c/1271778/ (Accessed: 22 November 2020).

FAO (2020b) *Smallholders and family farms in Armenia. Country Study Report*, *Smallholders and family farms in Armenia*. Budapest. doi: 10.4060/ca9823en.

Farr, T. G. *et al.* (2007) 'The shuttle radar topography mission', *Reviews Geophysics*, 45, pp. 1–33. doi: 10.1029/2005RG000183.

Fayvush, G., Aleksanyan, A. and Bussmann, R. W. (2016) 'Ethnobotany of the Caucasus – Armenia', pp. 1–17. doi: 10.1007/978-3-319-50009-6_18-1.

Fayvush, G. M. and Aleksanyan, A. S. (2020) 'The Transcaucasian Highlands', in Noroozi, J. (ed.) *Plant Biogeography and Vegetation of High Mountains of Central and South-West Asia*. Cham, Switzerland: Springer, pp. 287–313. doi: https://doi.org/10.1007/978-3-030-45212-4.

Feminó, M. (no date) *Beehives near Goris*, *Google Earth Photos*. (Accessed: 10 August 2020).

Foley, J. A. *et al.* (2005) 'Global consequences of land use', *Science*, 309, pp. 570–574. doi: 10.1126/science.1111772.

Foody, G. M. (2004) 'Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy', *Photogrammetric Engineering and Remote Sensing*, 70(5), pp. 627–633. doi: 10.14358/PERS.70.5.627.

Foody, G. M. and Mathur, A. (2004) 'Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification', *Remote Sensing of Environment*, 93(1–2), pp. 107–117. doi: 10.1016/j.rse.2004.06.017.

Franke, J., Keuck, V. and Siegert, F. (2012) 'Assessment of grassland use intensity by remote sensing to support conservation schemes', *Journal for Nature Conservation*, 20(3), pp. 125–134. doi: 10.1016/j.jnc.2012.02.001.

Fu, B. *et al.* (2017) 'Comparison of object-based and pixel-based Random Forest algorithm for wetland vegetation mapping using high spatial resolution GF-1 and SAR data', *Ecological Indicators*, 73, pp. 105–117. doi: 10.1016/j.ecolind.2016.09.029.

Gao, B.-C. (1996) 'NDWI A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space', *Remote Sens. Environ*, 58(3), pp. 257–266.

GEE (no date a) *Eigen Analysis*. Available at: https://developers.google.com/earth-engine/guides/arrays_eigen_analysis (Accessed: 5 January 2021).

GEE (no date b) *Sentinel-1 Algorithms | Google Earth Engine | Google Developers*. Available at: https://developers.google.com/earth-engine/guides/sentinel1 (Accessed: 12 November 2020).

GEE (no date c) *Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar Ground Range Detected, log scaling*. Available at: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD (Accessed: 12 November 2020).

Ghazaryan, H. G. (2013) *Brief Outline Of Soils In Armenia*. Proceeding the Economic Dimension of Land Degradation, Desertification and Increasing the Resilience of Affected Areas in the Region of Central and Eastern Europe (EDLDIR-2013). Brno Press, Czech Republik. ISBN 978-80-7375-715-1.

GIZ (no date) *ECOserve: Armenia*. Available at: https://biodivers-southcaucasus.org/countries/armenia (Accessed: 16 November 2020).

Gorbunov, V. (no date) *No Title*, *Google Earth Photos*. Available at: https://lh5.googleusercontent.com/p/AF1QipPynsVwoNzEanSJPyO2JFuCLirxD2Aj1-ya81yr=h1440 (Accessed: 10 August 2020).

Gorelick, N. *et al.* (2017) 'Google Earth Engine: Planetary-scale geospatial analysis for everyone', *Remote Sensing of Environment*, 202, pp. 18–27. doi: 10.1016/j.rse.2017.06.031.

Griffiths, P. *et al.* (2013) 'Agricultural land change in the Carpathian ecoregion after the breakdown of socialism and expansion of the European Union', *Environmental Research Letters*, 8(4). doi: 10.1088/1748-9326/8/4/045024.

Griffiths, P. *et al.* (2020) 'Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and Landsat time series', *Remote Sensing of Environment*, 238. doi: 10.1016/j.rse.2019.03.017.

Griffiths, P., Nendel, C. and Hostert, P. (2019) 'Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping', *Remote Sensing of Environment*, 220, pp. 135–151. doi: 10.1016/j.rse.2018.10.031.

Gulsoy, E. (2019) 'Determination of phenological and pomological properties and fatty acid contents of some wild almond genotypes (Prunus fenzliana fritsch) grown on the slopes of mount ararat', *Journal of Agricultural Science and Technology*, 21(6), pp. 1535–1546.

Gumma, M. K. *et al.* (2020) 'Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using random forest machine learning algorithms on the Google Earth Engine cloud', *GIScience and Remote Sensing*, 57(3), pp. 302–322. doi: 10.1080/15481603.2019.1690780.

Halabuk, A. *et al.* (2015) 'Towards detection of cutting in hay meadows by using of NDVI and EVI time series', *Remote Sensing*, 7(5), pp. 6107–6132. doi: 10.3390/rs70506107.

Hall-Beyer, M. (2017) 'Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales', *International Journal of Remote Sensing*, 38(5), pp. 1312–1338. doi: 10.1080/01431161.2016.1278314.

Hammond, T. O. and Verbyla, D. L. (1996) 'Optimistic bias in classification accuracy assessment', *International Journal of Remote Sensing*, 17(6), pp. 1261–1266. doi: 10.1080/01431169608949085.

Hartvigsen, M. (2014) 'Land reform and land fragmentation in Central and Eastern Europe', *Land Use Policy*, 36, pp. 330–341. doi: 10.1016/j.landusepol.2013.08.016.

Harutyunyan, A., Petrosyan, E., *et al.* (2019) 'Working Paper Energy Use and Demand of Communities in Getik', pp. 1–19. doi: 10.13140/RG.2.2.36734.13126.

Harutyunyan, A., Schlaffer, S., *et al.* (2019) 'Working Paper Land-Use and Land-Cover Mapping of Getik River Basin ', pp. 1–18.

Harutyunyan, A., Pfeiffer, P., *et al.* (2019) 'Working Paper: Public Participation GIS for Mapping Land Use Patterns in Getik Valley, Armenia', pp. 1–27. doi: 10.13140/RG.2.2.19956.91529.

Hast, I. and Mehari, A. (2016) *Automating Geographic Object-Based Image Analysis and Assessing the Methods Transferability: A Case Study Using High Resolution Geografiska Sverigedata Orthophotos*. University of Gävle. Available at: http://www.diva-portal.org/smash/get/diva2:1034429/FULLTEXT01.pdf. (Accessed: 14 September 2020).

Herher reservoir: low water season (2012) *Google Earth Photos*. Available at: https://lh5.googleusercontent.com/p/AF1QipNgasJFXRjhTTqVPZwg6_SzLkCS91CsnES9Kh-Z=h1440 (Accessed: 10 August 2020).

Herold, N. (2011) *Resolution Vs. Minimum Mapping Unit: Size Does Matter*. Available at: https://geozoneblog.wordpress.com/2011/10/28/resolution-vs-minimum-mapping-unit-size-does-matter/ (Accessed: 2 November 2020).

Hijmans, R. J. *et al.* (2020) 'Package "raster"', *Cran*, pp. 1–249. Available at: https://cran.r-project.org/web/packages/raster/raster.pdf.

Huete, A. R. *et al.* (1997) 'A comparison of vegetation indices over a global set of TM images for EOS-MODIS', *Remote Sensing of Environment*, 59(3), pp. 440–451. doi: 10.1016/S0034-4257(96)00112-5.

Hussain, E. and Shan, J. (2016) 'Object-based urban land cover classification using rule inheritance over very high-resolution multisensor and multitemporal data', *GIScience and Remote Sensing*, 53(2), pp. 164–182. doi: 10.1080/15481603.2015.1122923.

Jain, M. *et al.* (2013) 'Mapping cropping intensity of smallholder farms: A comparison of methods using multiple sensors', *Remote Sensing of Environment*, 134, pp. 210–223. doi: 10.1016/j.rse.2013.02.029.

Kauth, R. J. and Thomas, G. S. (1976) 'Tasselled Cap - a Graphic Description of the Spectral-Temporal Development of Agricultural Crops As Seen By Landsat.', pp. 41–51. The Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

Kolecka, N. *et al.* (2018) 'Regional scale mapping of grassland mowing frequency with Sentinel-2 time series', *Remote Sensing*, 10(8). doi: 10.3390/rs10081221.

Kraemer, R. *et al.* (2015) 'Long-term agricultural land-cover change and potential for cropland expansion in the former Virgin Lands area of Kazakhstan', *Environmental Research Letters*, 10(5). doi: 10.1088/1748-9326/10/5/054012.

Kuemmerle, T. *et al.* (2009) 'Land use change in Southern Romania after the collapse of socialism', *Regional Environmental Change*, 9(1), pp. 1–12. doi: 10.1007/s10113-008-0050-z.

Kuemmerle, T. *et al.* (2020) 'Identifying priority areas for restoring mountain ungulates in the Caucasus ecoregion', *Conservation Science and Practice*. doi: 10.1111/csp2.276.

Kumar, L. and Mutanga, O. (2018) 'Google Earth Engine applications since inception: Usage, trends, and potential', *Remote Sensing*, 10(10), pp. 1–15. doi: 10.3390/rs10101509.

Kuntz, C. (2019) *Community-based landscape conservation in Armenia*. Available at: https://panorama.solutions/en/solution/community-based-landscape-conservation-armenia (Accessed: 18 November 2020).

Lebourgeois, V. *et al.* (2017) 'A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM)', *Remote Sensing*, 9(3), pp. 1–20. doi: 10.3390/rs9030259.

Lomba, A. *et al.* (2014) 'Mapping and monitoring High Nature Value farmlands: Challenges in European landscapes', *Journal of Environmental Management*, 143, pp. 140–150. doi: 10.1016/j.jenvman.2014.04.029.

Lomba, A. *et al.* (2015) 'Reconciling nature conservation and traditional farming practices: A spatially explicit framework to assess the extent of High Nature Value farmlands in the European countryside', *Ecology and Evolution*, 5(5), pp. 1031–1044. doi: 10.1002/ece3.1415.

Lomba, A. *et al.* (2020) 'Back to the future: rethinking socioecological systems underlying high nature value farmlands', *Frontiers in Ecology and the Environment*, 18(1), pp. 36–42. doi: 10.1002/fee.2116.

Mahdianpari, M. *et al.* (2019) 'The first wetland inventory map of newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the Google Earth Engine cloud computing platform', *Remote Sensing*, 11(1). doi: 10.3390/rs11010043.

McFeeters, S. K. (1996) 'The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features', *International Journal of Remote Sensing*, 17(7), pp. 1425–1432. doi: 10.1080/01431169608948714.

McRoberts, R. E. (2011) 'Satellite image-based maps: Scientific inference or pretty pictures?', *Remote Sensing of Environment*, 115(2), pp. 715–724. doi: 10.1016/j.rse.2010.10.013.

Millard, K. and Richardson, M. (2015) 'On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping', *Remote Sensing*, 7(7), pp. 8489–8515. doi: 10.3390/rs70708489.

Millns, J. (2013) 'Agriculture and Rural Cooperation: Examples from Armenia, Georgia and Moldova', *Policy Studies on Rural Transition No. 2013-2*, pp. 1–37. Available at: http://www.dorada.org.ua/ar424e.pdf.

Moran, P. A. P. (1950) 'Notes on Continuous Stochastic Phenomena', *Biometrika*, 37(1/2), pp. 17–23. doi: 10.2307/2332142.

Muller, D. *et al.* (2009) 'Lost in transition: Determinants of post-socialist cropland abandonment in Romania', *Journal of Land Use Science*, 4(1–2), pp. 109–129. doi: 10.1080/17474230802645881.

Nanagulyan, S. *et al.* (2020) 'Wild plants and fungi sold in the markets of Yerevan (Armenia)', *Journal of Ethnobiology and Ethnomedicine*, 16(1), pp. 1–27. doi: 10.1186/s13002-020-00375-3.

NASA (no date) *Visible Infrared Imaging Radiometer Suite (VIIRS) - LAADS DAAC*. Available at: https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/viirs/ (Accessed: 12 November 2020).

do Nascimento Bendini, H. *et al.* (2019) 'Detailed agricultural land classification in the Brazilian cerrado based on phenological information from dense satellite image time series', *International Journal of Applied Earth Observation and Geoinformation*, 82. doi: 10.1016/j.jag.2019.05.005.

Numbisi, F. N., Van Coillie, F. and De Wulf, R. (2018) 'Multi-date sentinel1 SAR image textures discriminate perennial agroforests in a tropical forest-savannah transition landscape', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(1), pp. 339–346. doi: 10.5194/isprs-archives-XLII-1-339-2018.

Olofsson, P. *et al.* (2014) 'Good practices for estimating area and assessing accuracy of land change', *Remote Sensing of Environment*, 148, pp. 42–57. doi: 10.1016/j.rse.2014.02.015.

Pareeth, S. *et al.* (2019) 'remote sensing Mapping Agricultural Landuse Patterns from Time Series of Landsat 8 Using Random Forest Based Hierarchial Approach', 11(5). doi: 10.3390/rs11050601.

Peña, M. A. and Brenning, A. (2015) 'Assessing fruit-tree crop classification from Landsat-8 time series for the Maipo Valley, Chile', *Remote Sensing of Environment*, 171, pp. 234–244. doi: 10.1016/j.rse.2015.10.029.

Peña, M. A., Liao, R. and Brenning, A. (2017) 'Using spectrotemporal indices to improve the fruit-tree crop classification accuracy', *ISPRS Journal of Photogrammetry and Remote Sensing*, 128, pp. 158–169. doi: 10.1016/j.isprsjprs.2017.03.019.

Phalke, A. R. *et al.* (2020) 'Mapping croplands of Europe, Middle East, Russia, and Central Asia using Landsat, Random Forest, and Google Earth Engine', *ISPRS Journal of Photogrammetry and Remote Sensing*, 167, pp. 104–122. doi: 10.1016/j.isprsjprs.2020.06.022.

Planet (2016) *PLANET IMAGERY PRODUCT SPECIFICATION: PLANETSCOPE & RAPIDEYE*. Available at: tinyurl.com/vt5e34zh. (Accessed: 13 September 2020).

Plieninger, T. and Bieling, C. (2012) *Resilience and the Cultural Landscape: Understanding and Managing Change in Human-Shaped Environments*. Edited by T. Plieninger and C. Bieling. Cambridge University Press. doi: 10.1017/CBO9781139107778.

Prishchepov, A. V. *et al.* (2013) 'Determinants of agricultural land abandonment in post-Soviet European Russia', *Land Use Policy*, 30(1), pp. 873–884. doi: 10.1016/j.landusepol.2012.06.011.

Prishchepov, A. V. *et al.* (2012) 'The effect of satellite image dates on land cover change detection and the mapping of agricultural land abandonment in Eastern Europe', *Remote Sensing of Environment,* 126, pp. 195-209. doi: 10.1016/j.rse.2012.08.017.

Republic of Armenia (2018) 'SDG implementation- voluntary national review: Armenia', pp. 1–82. Available at: https://sustainabledevelopment.un.org/content/documents/20315 Armenia_SDG_VNR_report.pdf. (Accessed: 5 August 2020).

Richardson, A. J. and Wiegand, C. L. (1977) 'Distinguishing vegetation from soil background information', *Photogrammetric Engineering and Remote Sensing*, 43(12), pp. 1541–1552.

Rikimaru, A., Roy, P. S. and Miyatake, S. (2002) 'Tropical forest cover density mapping', *Tropical Ecology*, 43(1), pp. 39–47. ISSN 0564-3295.

Rouse, J. W. *et al.* (1973) 'Monitoring vegetation systems in the great plains with ERTS', *Third Earth Resources Technology Satellite (ERTS) symposium*, 1, pp. 309–317. Available at: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740022614.pdf.

Rozelle, S. and Swinnen, J. F. M. (2004) 'Success and failure of reform: Insights from the transition of agriculture', *Journal of Economic Literature*, 42(2), pp. 404–456. doi: 10.1257/0022051041409048.

Rufin, P. *et al.* (2019) 'Mapping cropping practices on a national scale using intra-annual Landsat time series binning', *Remote Sensing*, 11(3). doi: 10.3390/rs11030232.

Schlaffer, S. and Harutyunyan, A. (2018) *Remote Sensing of Land-Cover / Land-Use in the Voghji River Basin , Syunik Region, Armenia*. doi: 10.13140/RG.2.2.19942.65604.

Schulte, M. and Harutyunyan, A. (2020) *Drivers of Forest Cover Developments in Armenia : A Potential Afforestation Scenario of the Getik River Basin*. Project Paper. Available at: tinyurl.com/klu9bv12. (Accessed: 14 November 2020).

Senf, C. *et al.* (2015) 'Mapping land cover in complex Mediterranean landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery', *Remote Sensing of Environment*, 156, pp. 527–536. doi: 10.1016/j.rse.2014.10.018.

Shi, T. and Xu, H. (2019) 'Derivation of Tasseled Cap Transformation Coefficients for Sentinel-2 MSI At-Sensor Reflectance Data', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(10), pp. 4038–4048. doi: 10.1109/JSTARS.2019.2938388.

Spiker, J. S. and Warner, T. A. (2007) 'Scale and Spatial Autocorrelation From A Remote Sensing Perspective', in Jensen, R. R., Gatrell, J. D., and McLean, D. (eds) *Geo-Spatial Technologies in Urban Environments*. Berlin, Heidelberg: Springer. doi: https://doi.org/10.1007/978-3-540-69417-5_10.

Splinter, K. D., Harley, M. D. and Turner, I. L. (2018) 'Remote sensing is changing our view of the coast: Insights from 40 years of monitoring at Narrabeen-Collaroy, Australias', *Remote Sensing*, 10(11). doi: 10.3390/rs10111744.

SRTM (2015) 'The Shuttle Radar Topography Mission (SRTM) Collection User Guide', pp. 1–17. Available at: https://lpdaac.usgs.gov/documents/179/SRTM_User_Guide_V3.pdf. (Accessed: 14 September 2020).

Stefanski, J. *et al.* (2014) 'Mapping land management regimes in western Ukraine using optical and SAR data', *Remote Sensing*, 6(6), pp. 5279–5305. doi: 10.3390/rs6065279.

Stehman, S. V. (2000) 'Practical implications of design-based sampling inference for thematic map accuracy assessment', *Remote Sensing of Environment*. doi: 10.1016/S0034-4257(99)00090-5.

Stehman, S. V. (2001) 'Statistical rigor and practical utility in thematic map accuracy assessment', *Photogrammetric Engineering and Remote Sensing*, 67(6), pp. 727–734.

Sukhija, S., Krishnan, N. C. and Kumar, D. (2018) 'Supervised heterogeneous transfer learning using random forests', *ACM International Conference Proceeding Series*, pp. 157–166. doi: 10.1145/3152494.3152510.

Tamiminia, H. *et al.* (2020) 'Google Earth Engine for geo-big data applications: A meta-analysis and systematic review', *ISPRS Journal of Photogrammetry and Remote Sensing*, 164, pp. 152–170. doi: 10.1016/j.isprsjprs.2020.04.001.

Teluguntla, P. *et al.* (2018) 'A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform', *ISPRS Journal of Photogrammetry and Remote Sensing*, 144, pp. 325–340. doi: 10.1016/j.isprsjprs.2018.07.017.

Theissen, T. *et al.* (2019) 'Environmental and socio-economic resources at the landscape level – Potentials for sustainable land use in the Georgian Greater Caucasus', *Journal of Environmental Management*, 232, pp. 310–320. doi: 10.1016/j.jenvman.2018.11.024.

Tobler, W. R. (1970) 'A Computer Movie Simulating Urban Growth in the Detroit Region', *Economic Geography*, 46, pp. 234–240. doi: 10.2307/143141.

Van Tricht, K. *et al.* (2018) 'Synergistic use of radar sentinel-1 and optical sentinel-2 imagery for crop mapping: A case study for Belgium', *Remote Sensing*, 10(10), pp. 1–22. doi: 10.3390/rs10101642.

Urutyan, V. (2020) *Virtual Event - Discussion on the Implications of the 2020 Global Food Policy Report for Eurasia*, *Global Food Policy Report*. Available at: https://www.ifpri.org/event/virtual-event-discussion-implications-2020-global-food-policy-report-eurasia. (Accessed: 14 September 2020).

Valdiviezo-N, J. C. *et al.* (2018) 'Built-up index methods and their applications for urban extraction from Sentinel 2A satellite data: discussion', *Journal of the Optical Society of America A*, 35(1), p. 35. doi: 10.1364/josaa.35.000035.

Verma, D. and Jana, A. (2019) 'LULC classification methodology based on simple Convolutional Neural Network to map complex urban forms at finer scale: Evidence from Mumbai', *arXiv*, pp. 1–28. Available at: https://arxiv.org/ftp/arxiv/papers/1909/1909.09774.pdf (Accessed: 5 August 2020).

Waqar, M. M. *et al.* (2012) 'Development of New Indices for Extraction of Built-Up Area & Bare Soil', *Open Access Scientific Reports*, 1(1), pp. 1–4. doi: 10.4172/scientificreports.13.

Wilkinson, K. N. *et al.* (2012) 'Areni-1 Cave, Armenia: A chalcolithic - early bronze age settlement and ritual site in the southern Caucasus', *Journal of Field Archaeology*, 37(1), pp. 20–33. doi: 10.1179/0093469011Z.0000000002.

World Food Programme (2019) *Armenia country strategic plan (2019–2024)*. Rome. Available at: https://www.wfp.org/operations/am02-armenia-country-strategic-plan-2019-2024

Worldbank (2012) *The Republic of Armenia: Climate change and agriculture country note*. Washington, DC. Available at: http://hdl.handle.net/10986/27435.

WWF (no date) *Eco-corridors fund for the Caucasus. A partnership for living landscapes*. Available at: https://www.ecfcaucasus.org/about (Accessed: 18 November 2020).

Wymann von Dach, S. *et al.* (2013) *Mountain Farming is Family Farming: A contribution from mountain areas to the International Year of Family Farming 2014*. Rome: FAO, CDE, BOKU. doi: http://www.fao.org/docrep/019/i3480e/i3480e.pdf E-ISBN.

Xiong, J. *et al.* (2017) 'Automated cropland mapping of continental Africa using Google Earth Engine cloud computing', *ISPRS Journal of Photogrammetry and Remote Sensing*, 126, pp. 225–244. doi: 10.1016/j.isprsjprs.2017.01.019.

Xu, H. (2006) 'Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery', *International Journal of Remote Sensing*, 27(14), pp. 3025–3033. doi: 10.1080/01431160600589179.

Xu, H. (2008) 'A new index for delineating built-up land features in satellite imagery', *International Journal of Remote Sensing*, 29(14), pp. 4269–4276. doi: 10.1080/01431160802039957.

Xue, J. and Su, B. (2017) 'Significant remote sensing vegetation indices: A review of developments and applications', *Journal of Sensors*, 2017. doi: 10.1155/2017/1353691.

Yin, H. (2018) 'A tool for plotting NDVI, BSI (bare soil index) and Tasseled Cap transformation (Brightness, Greenness and Wetness) time series from Landsat SR and Sentinel 2 TOA time series; visual interpretation of cloud-free Landsat SR imagery.' provided by A. Prishchepov.

Yin, H. *et al.* (2018) 'Mapping agricultural land abandonment from spatial and temporal segmentation of Landsat time series', *Remote Sensing of Environment*, 210, pp. 12–24. doi: 10.1016/j.rse.2018.02.050.

Yin, H. *et al.* (2019) 'Agricultural abandonment and re-cultivation during and after the Chechen Wars in the northern Caucasus', *Global Environmental Change*, 55, pp. 149–159. doi: 10.1016/j.gloenvcha.2019.01.005.

Yuan, X. *et al.* (2019) 'Filtering the NPP-VIIRS nighttime light data for improved detection of settlements in Africa', *Remote Sensing*, 11(24), pp. 1–22. doi: 10.3390/rs11243002.

Zazanashvili, N., Manvelyan, K. and Heidelberg, A. (2020) *Leopard in the South Caucasus*: *Conservation Summary since 2000*. WW-F Caucasus, Tbilisi, Georgia. ISBN 978-9941-8-2318-3.

Zhang, M. *et al.* (2019) 'Mapping bamboo with regional phenological characteristics derived from dense Landsat time series using Google Earth Engine', *International Journal of Remote Sensing*, 40(24), pp. 9541–9555. doi: 10.1080/01431161.2019.1633702.

Zhang, R. *et al.* (2020) 'A novel feature-level fusion framework using optical and SAR remote sensing images for land use/land cover (LULC) classification in cloudy mountainous area', *Applied Sciences (Switzerland)*, 10(8), pp. 1–24. doi: 10.3390/APP10082928.

Zhang, W. *et al.* (2019) 'From woody cover to woody canopies: How Sentinel-1 and Sentinel-2 data advance the mapping of woody plants in savannas', *Remote Sensing of Environment*, 234, p. 111465. doi: 10.1016/j.rse.2019.111465.

Zhen, Z. *et al.* (2013) 'Impact of training and validation sample selection on classification accuracy and accuracy assessment when using reference polygons in object-based classification', *International Journal of Remote Sensing*, 34(19), pp. 6914–6930. doi: 10.1080/01431161.2013.810822.

Zhu, L. *et al.* (2018) 'A Review: Remote Sensing Sensors', *Multi-purposeful Application of Geospatial Data*. doi: 10.5772/intechopen.71049.

Appendix

Appendix I Normalized difference indices and spectral transformation with formula and description. The spectral derivates where calculated and used within the construction of the various RF input datasets.

Derivate name	*Formula*	*Description*
Normalized difference vegetation index (NDVI)	$NDVI = \frac{(nir - red)}{(nir + red)}$	Highly associated with vegetation content, canopy structure/LAI; for discrimination of healthy and dense vegetation; sensitive to shadow, edaphic and atmospheric effects; (Rouse *et al.*, 1973)
Enhanced vegetation index (EVI)	$EVI = 2.5 * [\frac{(nir - red)}{(nir + 6 * red - 7.5 * blue + 1)}]$	Vegetation Index which simultaneously corrects for atmospheric and edaphic effects (de-coupling from canopy background) (Huete *et al.*, 1997)
Normalized difference moisture index (NDMI)	$NDMI = \frac{nir - swir1}{nir + swir1}$	Associated with plant water stress; sensitive to vegetation water content; (Gao, 1996)
Modified (secondary) soil-adjusted vegetation index (MSAVI2)	$MSAVI2 = 0.5 * [(2nir + 1) - \sqrt{(2nir + 1)^2 - 8(nir - red)}$	Corrected for changes in the background caused by soil colour or surface soil moisture content; (Richardson and Wiegand, 1977)
Tasseled cap (TC) transformation (with adjusted coefficients for S2)	$Wetness = (blue * 0.2578) + (green * 0.2305) + (red * 0.0883) + (nir * 0.1071) + (swir1 * -0.7611) + (swir2 * -0.5308)$ $Brightness = (blue * 0.3510) + (green * 0.3813) + (red * 0.3437) + (nir * 0.7196) + (swir1 * 0.2396) + (swir2 * 0.1949)$ $Greeness = (blue * -0.3599) * (green * -0.3533) + (red * -0.4734) + (nir * 0.6633) + (swir1 * 0.0087) + (swir2 * -0.2856)$	Conversion of the original image bands into a new set of bands in a four-dimensional feature space; the TC components represent properties of terrestrial reflectance associated with a number of biophysical parameters, such as brightness with albedo, greenness with vegetation, and wetness with soil moisture; sensor-specific coefficients; (Kauth and Thomas, 1976; Shi and Xu, 2019)
Normalized difference water index (NDWI)	$NDWI = \frac{green - nir}{green + nir}$	Delineation of open water features (McFeeters, 1996)
Modified normalized difference water index (MNDWI)	$MNDWI = \frac{green - swir1}{green + swir1}$	Delineation of open water features with correction for noise of land and urban features and structures (Xu, 2006)
Bare soil index (BSI) (adjusted)	$BSI = (\frac{(red + swir2) - (blue + nir)}{(red + swir2) - (blue + nir)})$	Associated with bare soil; highlights difference between agricultural and non-agricultural land; originally with swir1; adjusted to swir2; (Rikimaru, Roy and Miyatake, 2002; Diek *et al.*, 2017)
Normalized difference built-up area index (NBAI)	$NBAI = \frac{swir2 - swir1/green}{swir2 + swir1/green}$	Separation of bare and urban areas; (Waqar *et al.*, 2012; Valdiviezo-N *et al.*, 2018)
Index-based built-up index (IBI)	$IBI = \frac{[NDBI - SAVI + MNDWI/2]}{[NDBI + SAVI + MNDWI/2]}$	Major urban components of vegetation, water and built-up land for urban feature extraction (Xu, 2008)

Appendix II Detailed description of the collection framework for the training data.

The GEE sample collection framework has been constructed primarily with Sentinel-2 MSI Level-2A data. Besides cloud free multi-temporal median composites (2 month) of the phenological period of 2019, a greenest pixel quality composite based on a derived NDVI time series of the entire year was included. The median composite time series was displayed as true- and false-color composites. Due to spectral differences between two consecutive composites a detection of management signals such as ploughing or mowing was possible in order to identify associated classes such as field crops or hay meadows. The greenest pixel quality composite consists of the max NDVI values per pixel for a given time period (here: one year). It was visualized with a color palette from yellow to green with low max NDVI's displayed in yellow and high values displayed in green. The greenest pixel composite was useful in order to discriminate areas with bare ground or sparse vegetation. A pansharpened summer composite, elevation zones following the altitudinal zonation and unsupervised cluster maps of a median composite for summer and a bi-seasonal layerstack of a median composite for both summer and spring complemented the framework. The unsupervised cluster results were barely used but gave a good impression about spectral similarity. The elevation zones helped to rule out some classes and the pansharpened image enhanced the visual impression of e.g. urban structures.

The time series framework is based on a prepared GEE script (Yin, 2018). It is thought of as a tool for plotting NDVI, BSI and Tasseled Cap transformation (brightness, greenness and wetness) time series from Landsat surface reflectance and Sentinel 2 top-of-atmosphere as well as for visual interpretation of cloud-free Landsat surface reflectance imagery. Besides the time series plots for the mentioned derivates the script creates four windows for visual observations with: (1) Google Earth imagery, (2) the least clouded imagery for an early phenological period (i.e. spring) (3) the least clouded imagery for a late phenological period (i.e. summer) and (4) the least clouded imagery for the whole phenological period. The start and end dates for the different periods can be individually set by the user. The time series charts in Appendix III were generated using this framework.

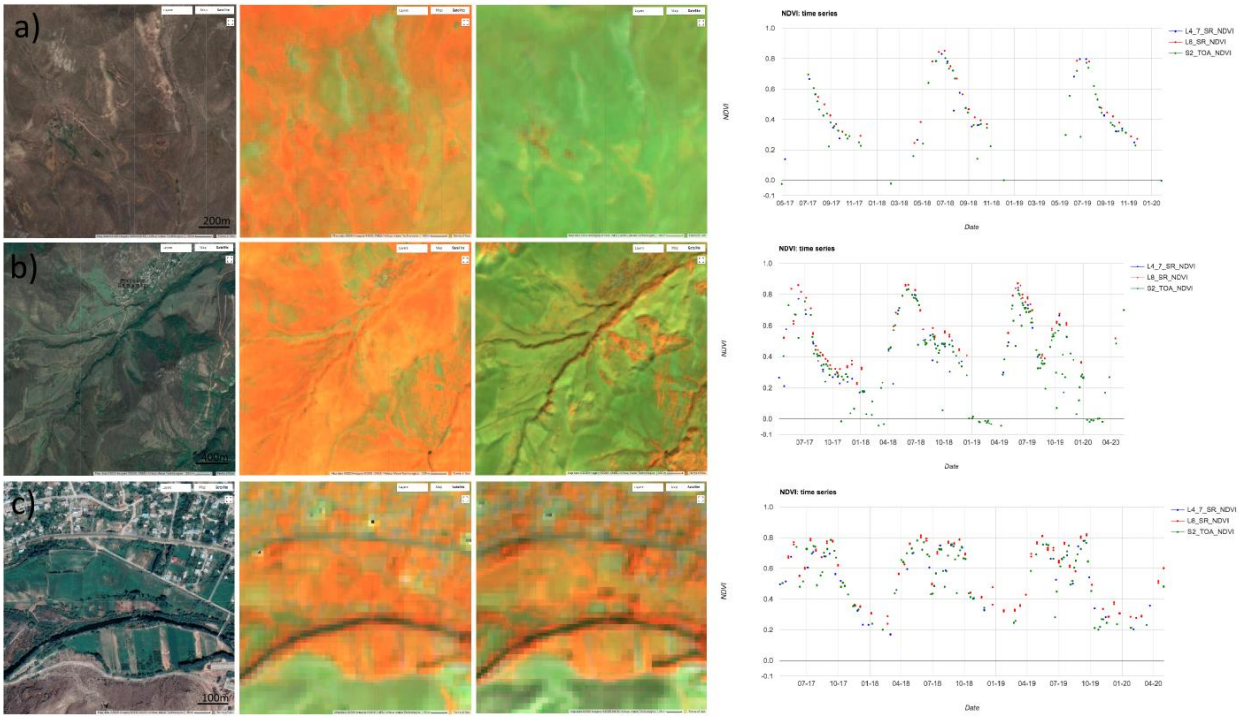
Appendix III illustrates the agricultural target classes with components of the sampling framework with: (1) a Google Earth VHRS composite (2019), (2) two Sentinel-2 seasonal composites for spring and summer and (3) a time series plot based on the NDVI of Landsat and

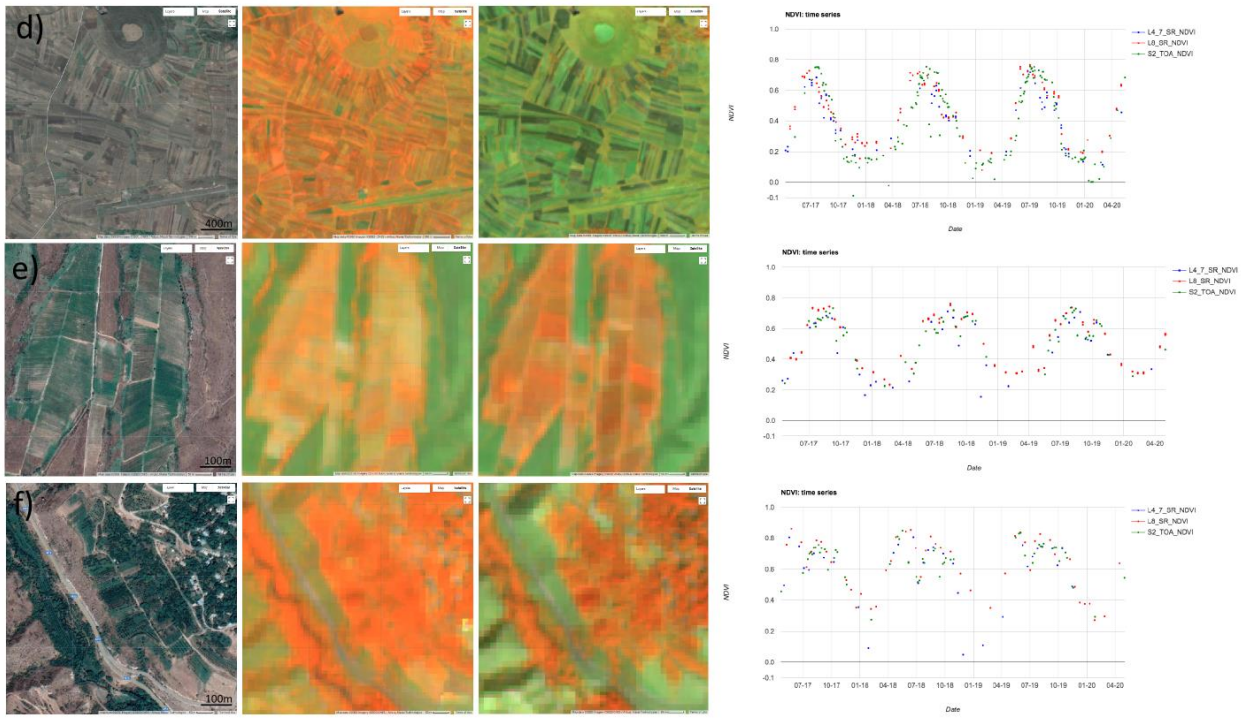
Sentinel-2. The grassland area is situated in the subalpine zone and shows a livestock station for the summer used by the transhumance pastoralists close to Jermuk in Vayots Dzor (Appendix III, a). The difference between the false color composite of spring and summer displays and even brightening due to the natural senescence of the herbaceous vegetation. Because of the altitude the phenological period starts quite late (May-June). In general, it shows a harmonic behavior with a greening phase maximum values around July and a ripening phase. It does not display a harvest signal and minimum values are above a NDVI of 0.2. The area for hay meadows is situated in a subalpine valley close to Tatev in Syunik (Appendix III, b). In the VHSR image of Google Earth harvested or mowed areas are visible. This is also displayed by the difference of the seasonal Sentinel-2 composites for spring and summer. Here the brightening is stronger for the harvested areas. Since no ploughing signal was detected for most of the harvested fields it seems to be mowed meadows. This is confirmed by a close look at the time series. It shows relatively high minimum values and no ploughing signal. The presence of a mowing signal is shown by a disturbed senescence. The area of improved meadows i.e. meadows which are irrigated and mowed more than once is located in the Arpa valley close to Yeghegnadzor (Appendix III, c). The VHSR and the seasonal false color composites display the persistent greenness of the meadows. This can be also observed by looking at the time series exhibiting consistently high NDVI values with some sudden drops associated with multiple mowing events.

Area d) in Appendix III shows crop fields close to Goris. The field crops show both a harvesting and a ploughing signal that can be observed by inspecting both the seasonal false color composites and the time series charts. Field crops exhibit very low minimum NDVI values and a ploughing signal can be quite pronounced. The site for vine yards is also located near Yeghegnadzor and is part of an extensive grape production area (Appendix III, e). Vine yards were mainly identified by the row structure. However, irrigated vine yards also exhibit a strong grass signal with high NDVI values over the whole phenological period. The grass signal is accompanied with the woody vegetation signal from the vines. The exemplary area of fruit orchards is situated in an arid mountain valley close to the Yeghegis state sanctuary (Appendix III, f). The more or less systematic structure of the orchards can be observed on the Google Earth VHRS. Within the more arid parts of the study region (Vayots Dzor) the orchards are also irrigated. As a function of the orchard density the tree signal is associated with a grass signal which may also show signs of

mowing management. Due to the high spatial and temporal resolution the VHSR imagery of the Planet Lab archive was in addition very useful for the detection of ploughing and mowing signals. Appendix IV illustrates the visual signal for both management practices. In combination with the time series framework and the GEE framework agricultural management could be identified successfully, even for small patch sizes.

Appendix III Illustration of the GEE training sample collection framework.





Appendix IV Visual signals as observed from the VHSR imagery in Planet explorer. a) ploughing signal and b) mowing signal.

Appendix V Acquired user's accuracy estimates of different publications. The estimates serve as broad indication of likely class specific accuracies for the estimation of validation sample sizes with the FAO spreadsheet.

LULC Class	*User's Accuracies*	*Publications*
Grassland/ potential pastures	88%; 50%;76%; 84.7%	(Das and Das, 2014); (Prishchepov *et al.*, 2012); (Schlaffer and Harutyunyan, 2018) (Senf *et al.*, 2015);
Hay meadows	60%; 74-88%; 57-69%[2]; 95%; 73%-91%; 77%	(Dusseux *et al.*, 2014); (Estel *et al.*, 2018); (Franke, Keuck and Siegert, 2012); (Prishchepov *et al.*, 2012); (Halabuk *et al.*, 2015); (Kolecka *et al.*, 2018);
Improved meadows	87.2%[3]	(Pareeth *et al.*, 2019)
Field crops	80%; 53-100%; 72/88%; 93.8%; 43%; 88.4%	(Das and Das, 2014); (Phalke *et al.*, 2020); (Pareeth *et al.*, 2019); (Prishchepov *et al.*, 2012); (Schlaffer and Harutyunyan, 2018); (Senf *et al.*, 2015)
Vine yards	85%	(Griffiths, Nendel and Hostert, 2019)
Fruit orchards	70-100%; 95%[4]; 88.6%	(Brinkhoff, Vardanega and Robson, 2020); (Das and Das, 2014); (Pareeth *et al.*, 2019)
Shrubland	50%; 50%; 43%	(Prishchepov *et al.*, 2012); (Senf *et al.*, 2015); (Schlaffer and Harutyunyan, 2018)
Forest	75%; 96.1%; 88%; 66%	(Griffiths, Nendel and Hostert, 2019); (Prishchepov *et al.*, 2012); (Schlaffer and Harutyunyan, 2018) ; (Senf *et al.*, 2015)
Bare/ sparse vegetation	95.8%/ 88.2%; 63%; 65%	(Pareeth *et al.*, 2019); (Schlaffer and Harutyunyan, 2018); (Senf *et al.*, 2015)
Water	85%; 85%	(Griffiths, Nendel and Hostert, 2019); (Schlaffer and Harutyunyan, 2018)
Urban	75-80%; 96%; 86%	(Griffiths, Nendel and Hostert, 2019); (Pareeth *et al.*, 2019); (Schlaffer and Harutyunyan, 2018)

[2] For different mowing intensity classes.

[3] Irrigated agriculture class used for imroved irrigated meadows.

[4] For homegardens.

Appendix VI Different validation sample allocations with varying optimism regarding the accuracy estimates obtained from the review (see Appendix V). U is the assumed user's accuracy of each class and n is the required sample size.

LULC Class	*Alloc1*		*Alloc2*		*Alloc3*	
	U_i	n_i	U_i	n_i	U_i	n_i
Grassland/ potential pastures	0,90	680,25	0,80	1238,43	0,70	1522,70
Hay meadows	0,85	43,41	0,80	79,03	0,75	97,17
Improved meadows	0,80	2,80	0,70	5,09	0,80	6,26
Field crops	0,80	13,06	0,80	23,78	0,80	29,24
Vine yards	0,80	7,80	0,80	14,20	0,80	17,46
Fruit orchards	0,80	14,76	0,60	26,87	0,60	33,04
Shrubland	0,90	123,99	0,50	225,72	0,50	277,53
Forest	0,90	15,04	0,90	27,38	0,80	33,66
Bare/ sparse vegetation	0,90	31,33	0,80	57,03	0,80	70,13
Water	0,90	1,41	0,90	2,57	0,85	3,17
Urban	0,95	4,56	0,70	8,31	0,80	10,21
Total		938,41		1708,42		2100,57

Appendix VII Advocated allocations of validation samples per class strata, the assumed user's accuracies and the associated standard errors for both sub-regions.

		Vayots Dzor		Sisisan-Goris	
LULC Class	*Assumed UA*	*Sample size*	*Standard error*	*Sample size*	*Standard error*
Grassland/ potential pastures	90%	690	1.14%	660	1.97%
Hay meadows	80%	200	2.84%	200	2.84%
Improved meadows	70%	150	3.75%	150	3.75%
Field crops	80%	200	2.84%	200	2.84%
Vine yards	80%	200	2.84%	n.d.	n.d.
Fruit orchards	60%	200	3.47%	200	3.47%
Shrubland	50%	200	3.54%	150	4.1%
Forest	90%	70	3.61%	80	3.38%
Bare/ sparse vegetation	80%	80	4.5%	90	4.24%
Water	90%	50	4.29%	60	3.91%
Urban	70%	60	5.96%	70	5,52%
Total		2100		1860	

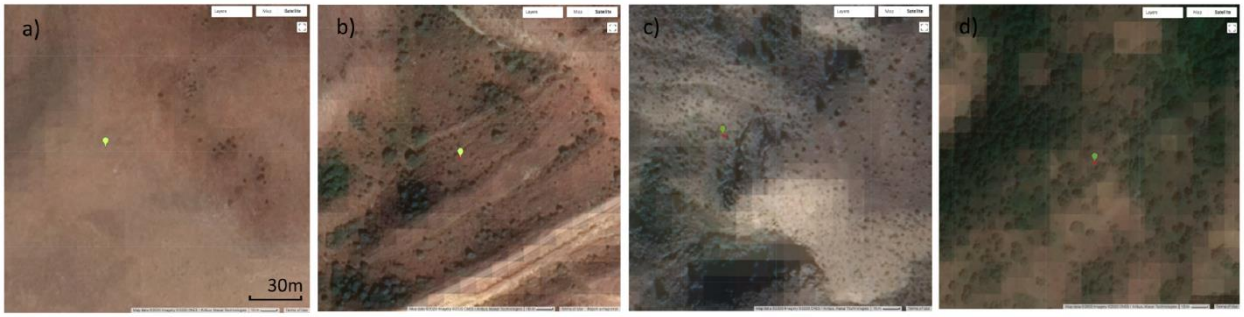
Appendix VIII Descriptions of easy and difficult cases during the flagging process of the response design within the validation sampling protocol.

Easy decisions have been the case for the forest class where the assessment units have been largely homogeneous. However, already for water and urban also some difficult cases have been faced regarding mixed pixels. Further, difficulties emerged for cases where the unit was not unambiguously a member of one of the defined discrete class options but rather in between of two classes. This is because land-cover is a continuum with transition zones and the representation of this continuum is generally done on a discrete scale (see Appendix XVIII). The

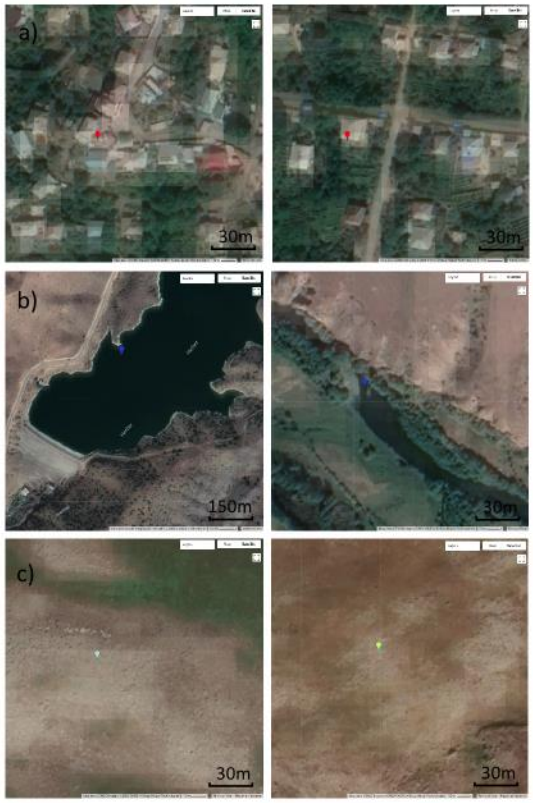
discrete class issue has been in particular the case for the classes bare and sparse vegetation, grassland and shrubland. In these cases, a definition of value thresholds and visual thresholds (i.e. proportions of pixel content) was necessary in order to agree on a consistent labeling process.

In the case of bare surfaces and sparse vegetation a greenest pixel quality composite based on a Sentinel2 NDVI time series for the year 2019 served as threshold source. Here a maximum NDVI threshold of 0.35 and 0.2 have been determined for Sisian-Goris and Vayots Dzor respectively (see Appendix XIX). The threshold where defined prior to the training sampling and have been derived from an initial assessment of greenest pixel values in different bare and sparsely vegetated areas. For the continuum of grass and shrubland and for other discrete classes such as urban or water coverage threshold of the assessment unit have been defined expressed by the coverage proportion. In the case of urban fabric, the threshold was below 50% due to the intensive spectral signal of this class (see Appendix XIX). Water was assigned with a threshold around 50% coverage of the MMU. The grassland and shrubland continuum was allocated with unit coverage threshold of 60-70% (see Appendix XIX).

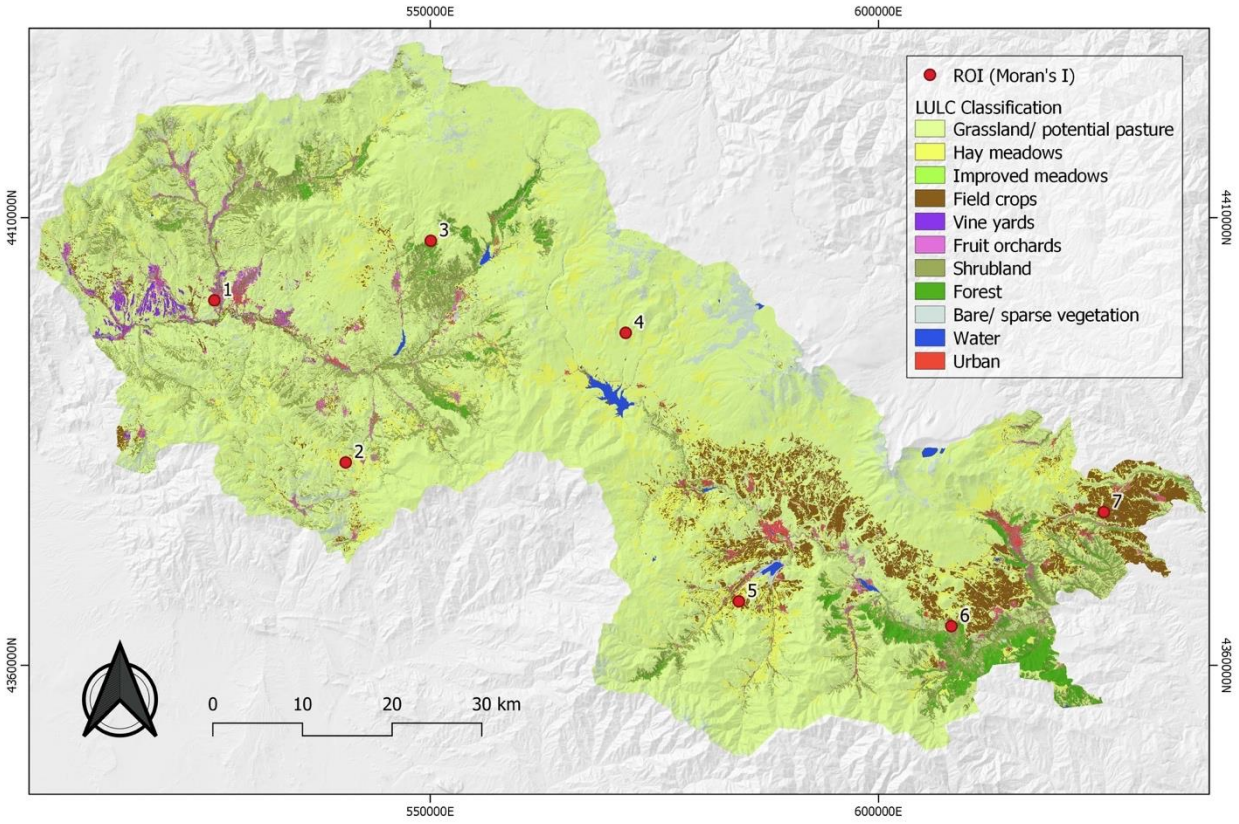
Appendix IX Grassland shrubland continuum.



Appendix X Ambiguous cases of for the urban (a), the water (b) and the bare class (c).



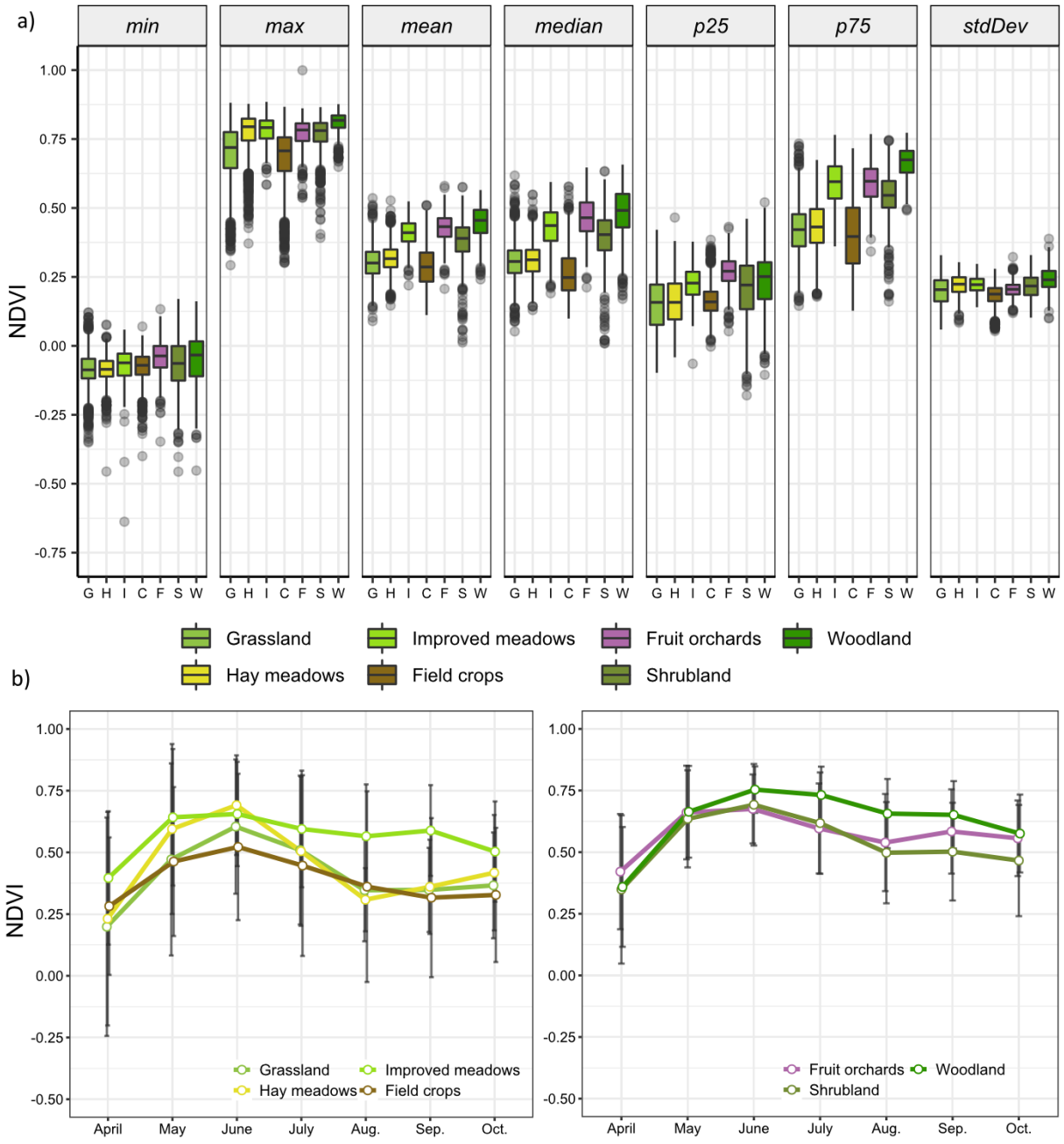
Appendix XI Classified map illustrating seven different landscapes considered for the calculation of Moran's I marked with points. The sub-sets exhibit distinct environments with varying land-use and land cover types found in the mountainous area. Some of the areas exhibit a more heterogenous distribution of classes (ROI 2, 5 & 6) whereas other parts show the presence of large clusters of agricultural fields (ROI 7) or natural grassland (ROI 4). In the mountain valleys of Vayots Dzor and in settlements clusters of vine yards and fruit orchards are apparent (ROI 1) (source: own illustration).



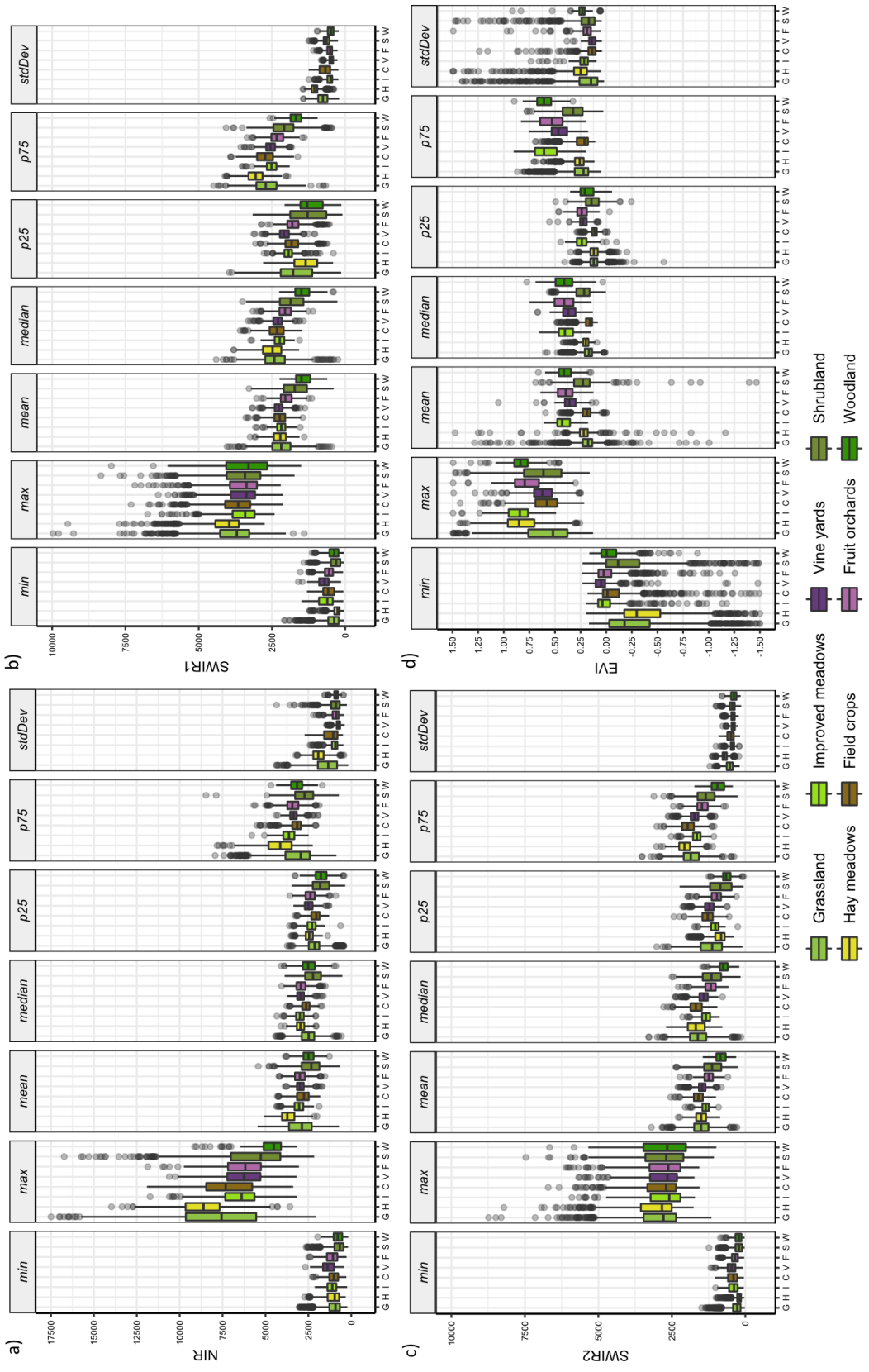
Appendix XII Example of a basic 4x4 error matrix with four different classes (Olofsson *et al.*, 2014).

		Reference				
		Class 1	*Class 2*	*Class 3*	*Class 4*	*Total*
Map	*Class 1*	p_{11}	p_{12}	p_{13}	p_{14}	$p_{1\cdot}$
	Class 2	p_{21}	p_{22}	p_{23}	p_{24}	$p_{2\cdot}$
	Class 3	p_{31}	p_{32}	p_{33}	p_{34}	$p_{3\cdot}$
	Class 4	p_{41}	p_{42}	p_{43}	p_{44}	$p_{4\cdot}$
	Total	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$	$p_{\cdot 4}$	1

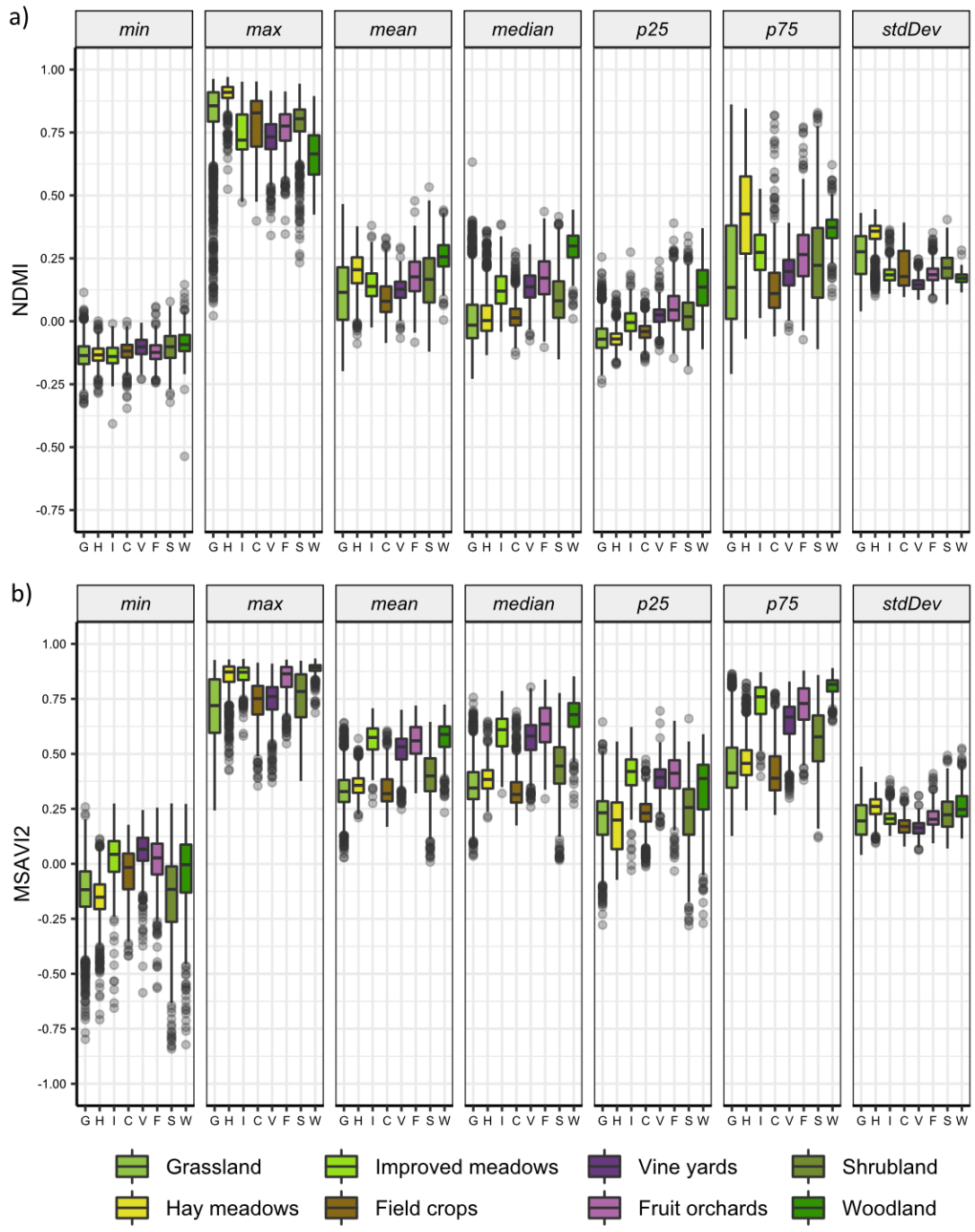
Appendix XIII Knowledge capture of the spectral-temporal NDVI characteristics of the agricultural and semi-natural LULC classes for Sisian-Goris. (a) the box- and whisker-plots for the spectral-temporal metrics, and (b) the monthly time series as line chart. The outliers in the box-plots are depicted as grey dots.



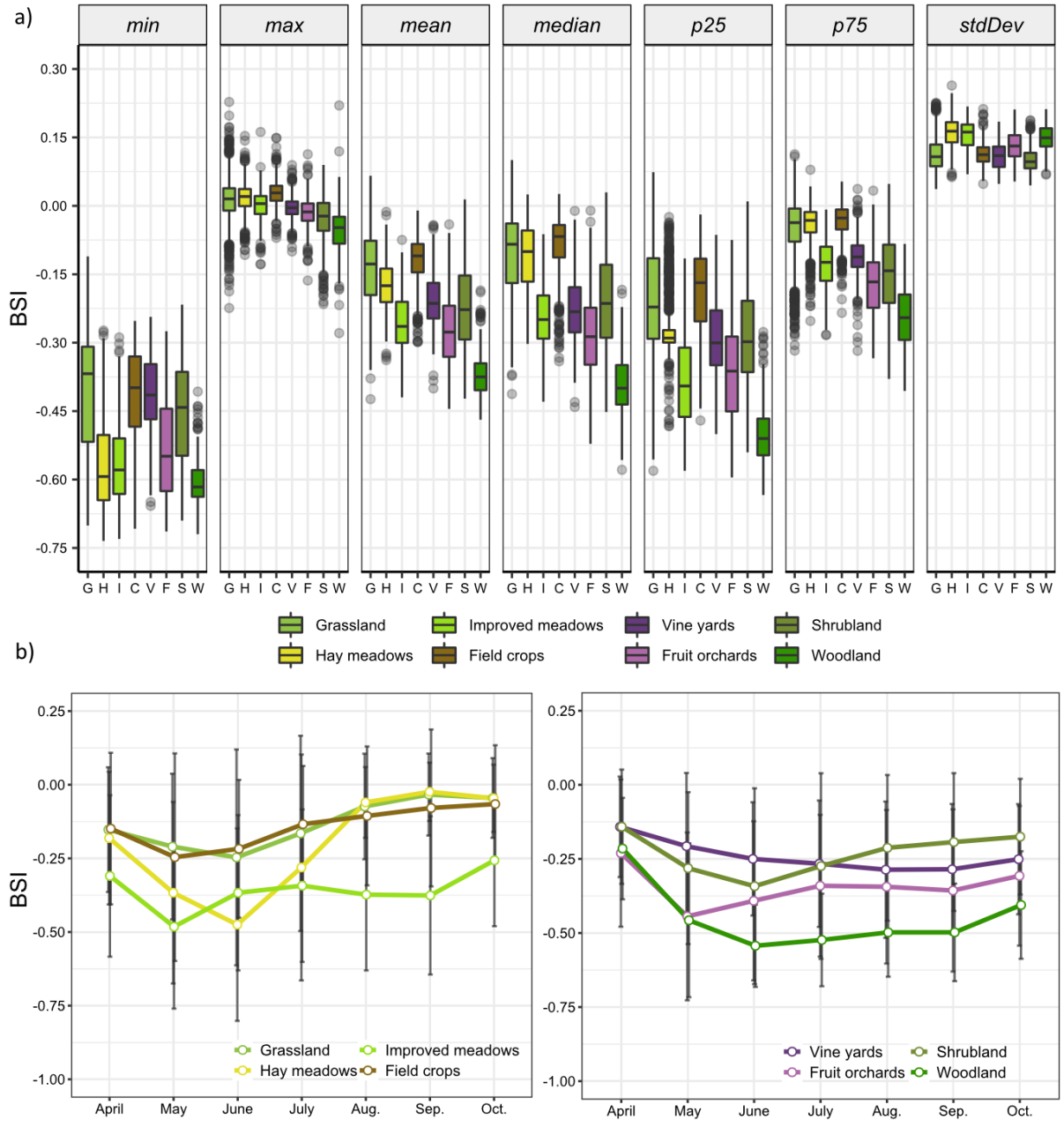
Appendix XIV Knowledge capture of the spectral-temporal characteristics of the agricultural and semi-natural LULC classes for Vayots Dzor. With spectral-temporal metrics of (a) the NIR band; (b) the SWIR1 band; (c) the SWIR2 band; and (d) the EVI.



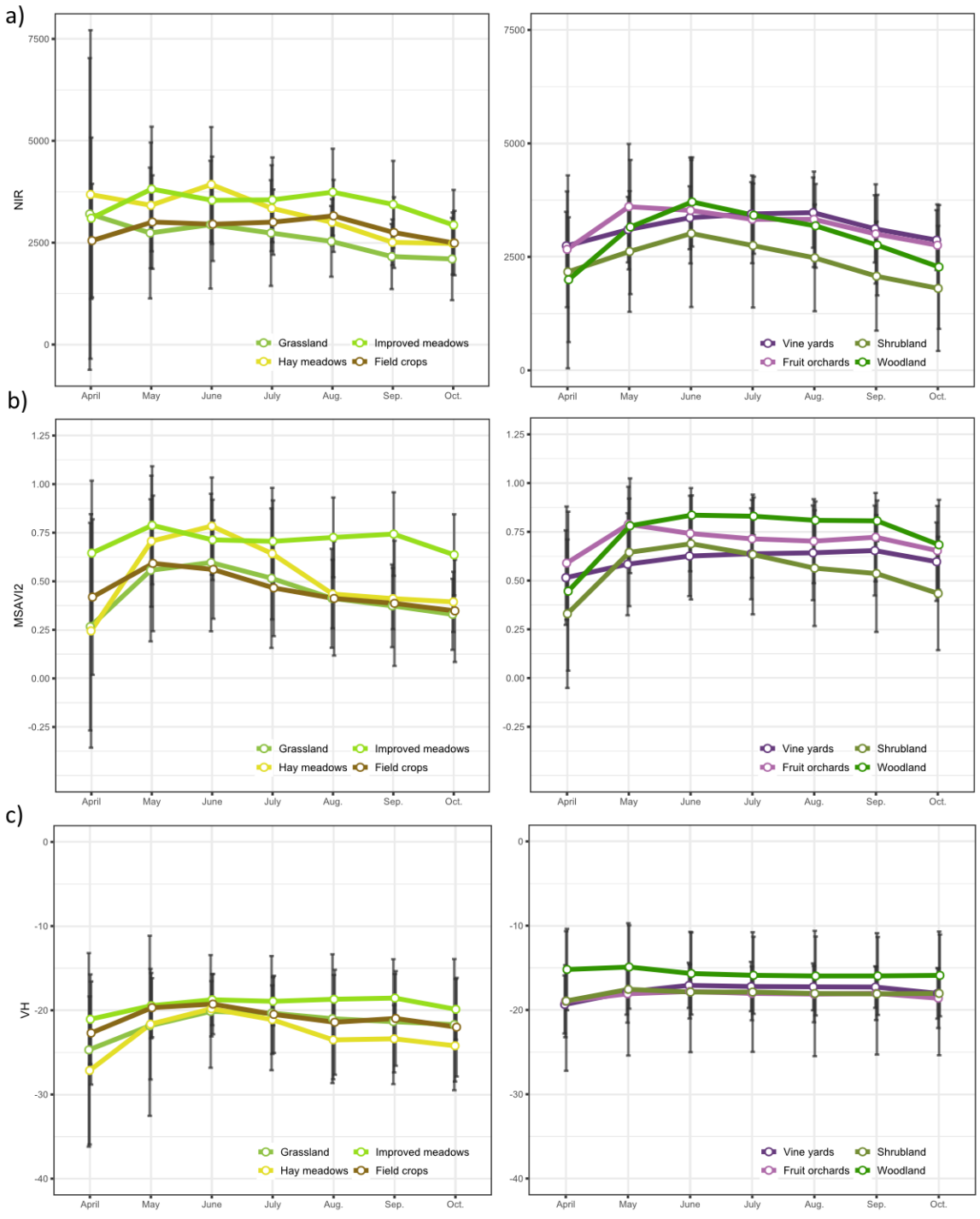
Appendix XV Knowledge capture of the spectral-temporal characteristics of the agricultural and semi-natural LULC classes for Vayots Dzor. With spectral-temporal metrics of (a) the NDMI; and (b) the MSAVI2.



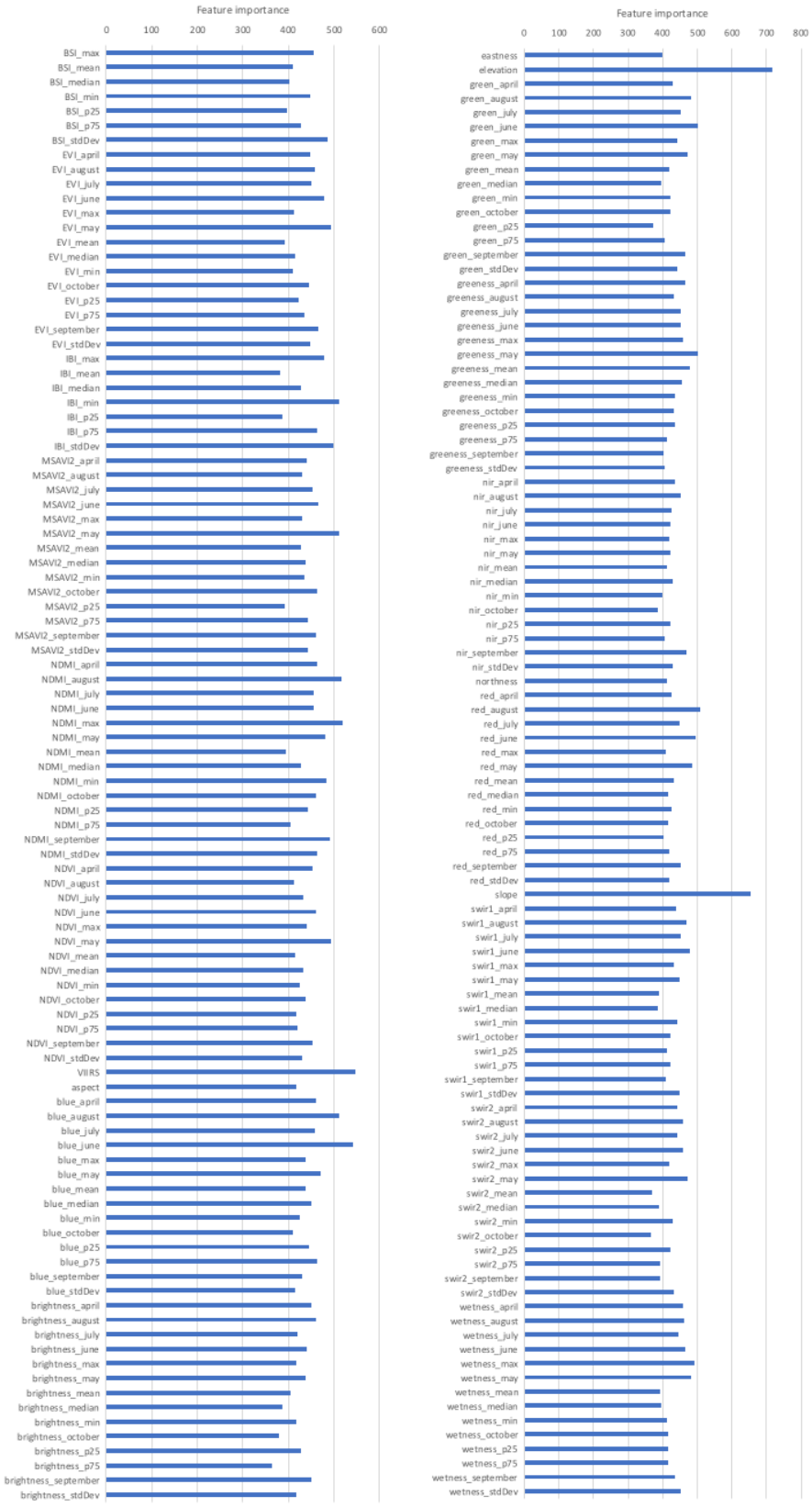
Appendix XVI Knowledge capture of the spectral-temporal BSI characteristics of the agricultural and semi-natural LULC classes for Sisian-Goris. (a) the box- and whisker-plots for the spectral-temporal metrics, and (b) the monthly time series as line chart.



Appendix XVII Time series of the NIR band, the MSAVI2 and the VH composites for the agricultural and semi-natural classes in Vayots Dzor.



Appendix XVIII RF feature importance of the "good choice" classification for Vayots Dzor.



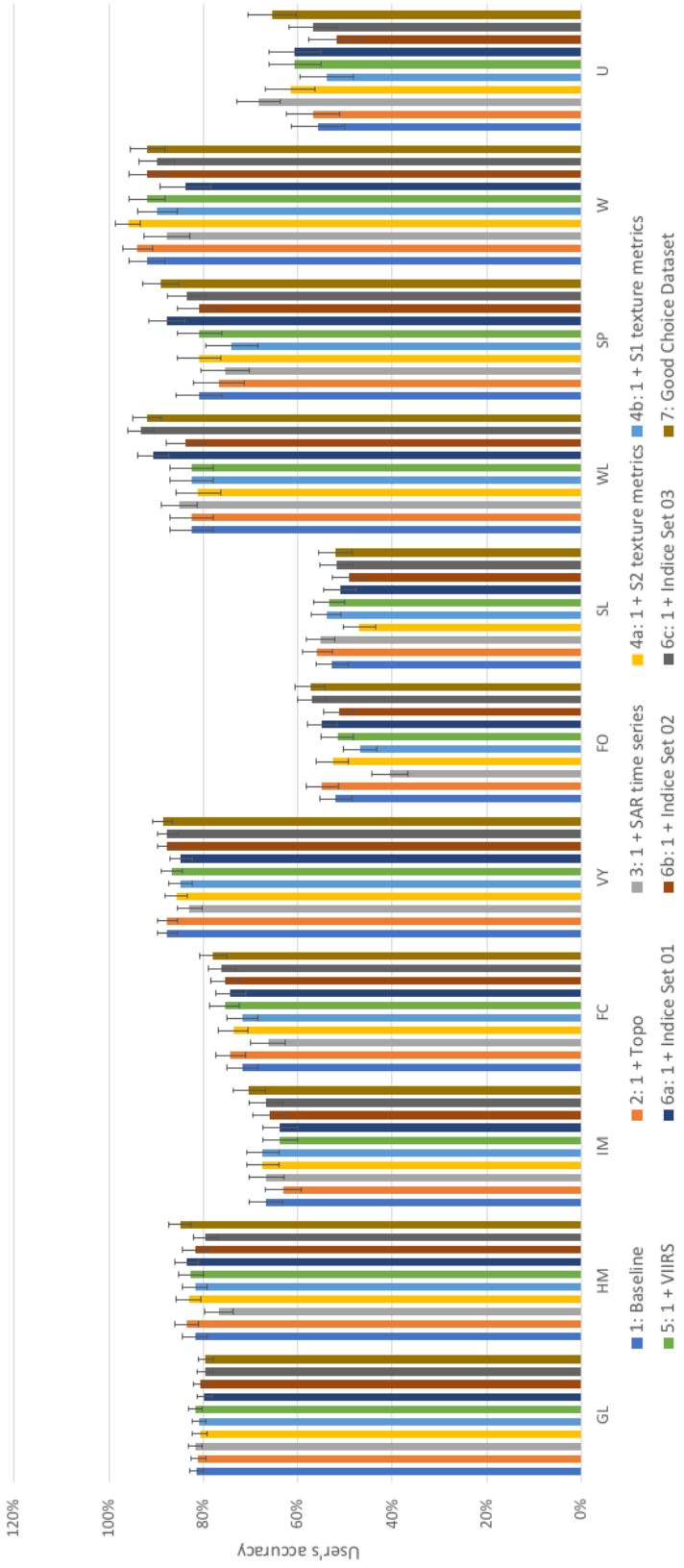
Appendix XIX Sample count-based error matrix with area-adjusted accuracy estimates of the baseline classification for the 11 classes of Vayots Dzor

		Reference													
		GL	HM	IM	FC	VY	FO	SL	WL	SP	W	U	∑	UA (%)	±
Map	GL	563	20	5	32	14	8	34	0	4	1	10	691	81,48	1,49
	HM	27	179	2	9	0	1	1	0	0	0	0	219	81,74	2,68
	IM	2	4	92	1	24	12	3	0	0	0	0	138	66,67	3,56
	FC	15	3	10	117	15	2	0	0	1	0	0	163	71,78	3,27
	VY	0	0	3	2	92	5	3	0	0	0	0	105	87,62	2,16
	FO	2	1	54	9	54	144	9	2	0	0	2	277	51,99	3,43
	SL	62	1	10	11	19	28	165	7	1	5	4	313	52,72	3,31
	WL	0	0	0	0	0	5	7	61	0	0	1	74	82,43	4,58
	SP	4	1	0	3	0	0	0	0	59	2	4	73	80,82	4,81
	W	1	0	0	0	0	1	2	0	0	45	0	49	91,84	3,76
	U	9	0	0	6	15	7	5	0	3	1	58	104	55,77	5,62
	∑	685	209	176	190	233	213	229	70	68	54	79	2206		
	PA (%)	94,28	66,59	14,40	20,19	22,51	24,73	63,59	81,41	84,39	26,81	17,59		OA (%)	±
	±	0,03	0,60	2,15	0,68	1,22	2,00	0,45	3,01	0,83	18,18	2,79		77,15	1,18

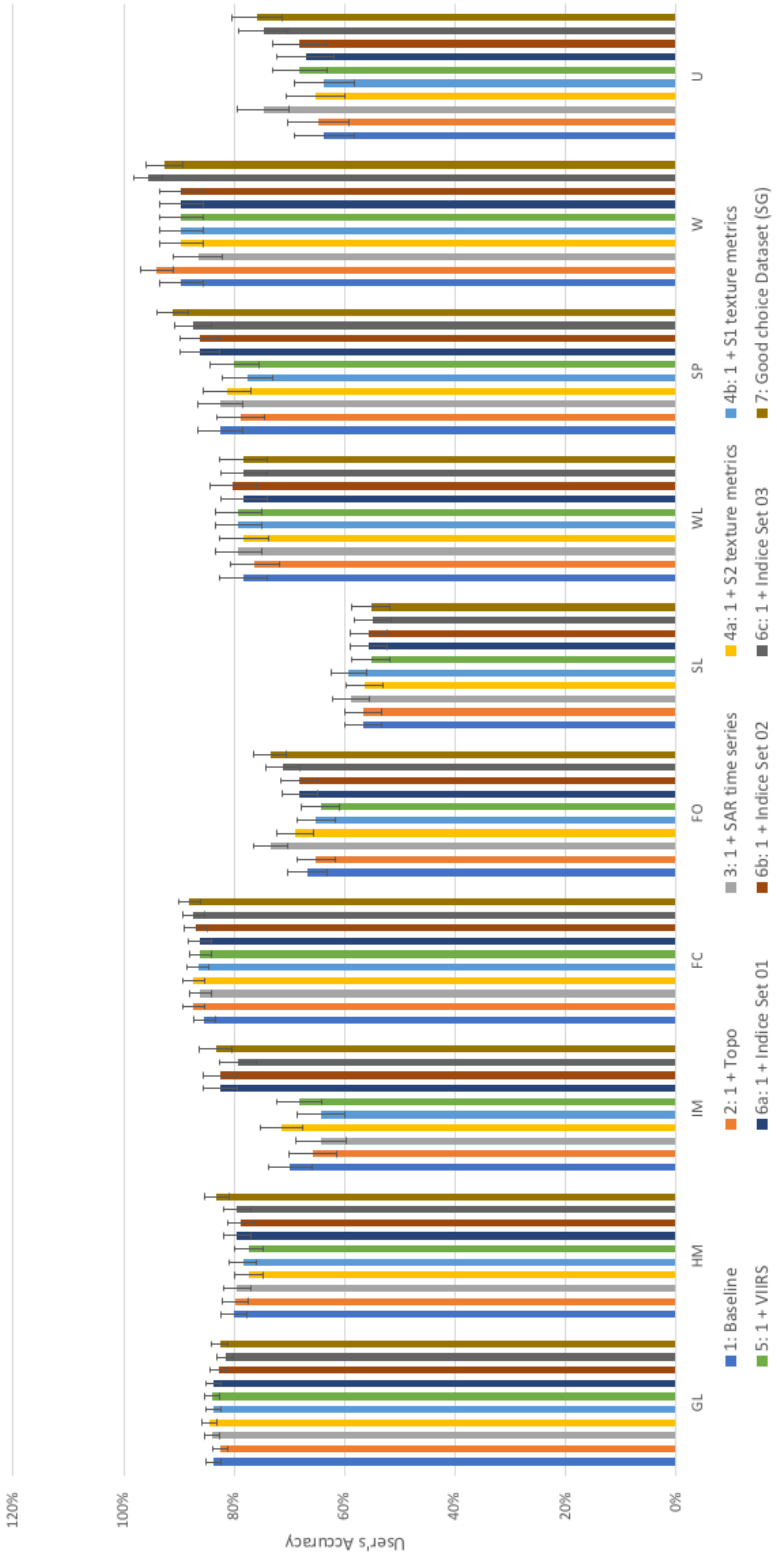
Appendix XX Sample count-based error matrix of baseline classification with accuracy estimates and associated standard errors for Sisian-Goris.

		GL	HM	IM	FC	FO	SL	WL	SP	W	U	∑	UA (%)	±
Map	GL	656	31	5	21	21	22	0	12	2	13	783	83,78	1,34
	HM	36	234	5	14	2	1	0	0	0	0	292	80,14	2,36
	IM	12	6	88	12	5	3	0	0	0	0	126	69,84	4,01
	FC	17	12	2	216	2	1	0	3	0	0	253	85,38	2,05
	FO	2	3	17	5	88	12	4	0	0	1	132	66,67	3,54
	SL	29	1	15	12	47	156	10	1	1	3	275	56,73	3,35
	WL	0	0	0	0	4	18	79	0	0	0	101	78,22	4,30
	SP	3	0	0	4	0	1	0	66	0	6	80	82,50	4,05
	W	1	0	0	2	1	3	0	0	60	0	67	89,55	3,88
	U	3	0	0	12	8	3	0	7	0	58	91	63,74	5,38
	∑	759	287	132	298	178	220	93	89	63	81	2200		
	PA (%)	95,55	69,75	8,15	74,83	14,69	55,23	91,76	71,60	72,21	24,17		OA (%)	±
	±	0,03	0,35	1,87	0,50	0,86	1,42	1,03	0,96	4,52	3,28		81,51	0,98

Appendix XXI Bar chart illustrating the differences of the user's accuracies between the different RF models for Vayots Dzor and all LULC classes. The standard error estimates are added as whiskers.



Appendix XXII Bar chart illustrating the differences of the user's accuracies between the different RF models for Sisian-Goris and all LULC classes. The standard error estimates are added as whiskers.



,