Université catholique de Louvain
Faculté des Bioingénieurs
Earth and Life Institute

Measuring Grassland Use Intensity by Remote Sensing for Agroecological Monitoring

---

Mathilde De Vroey

Thèse présentée en vue de l'obtention du grade de

Docteur en sciences agronomiques et ingénierie biologique

Thesis committee:

Pr. Pierre Defourny (Co-supervisor)	UCLouvain, Belgium
Dr. Julien Radoux (Co-supervisor)	UCLouvain, Belgium
Pr. Patrick Meyfroidt	UCLouvain, Belgium
Pr. Marc Dufrêne	ULiège, Belgium
Dr. Benjamin Koetz	European Space Agency, Italy
Pr. Emmanuel Hanert (Chair)	UCLouvain, Belgium

Louvain-la-Neuve, Janvier 2023

Diversity creates harmony,
and harmony creates beauty, balance, bounty and peace
in nature and society, in agriculture and culture,
in science and in politics.

Vandana Shiva

Acknowledgements - Remerciements

Cette thèse est le résultat d'un peu plus de quatre années de recherche. Bien que la réalisation d'une thèse de Doctorat soit une épreuve individuelle, un bon nombre de personnes ont directement ou indirectement contribué à ce travail.

Premièrement, j'aimerais particulièrement remercier mes co-promoteurs, Pierre Defourny et Julien Radoux pour leur encadrement très complémentaire, sans lequel cette thèse n'aurait pas pu être menée à bout. Merci à Pierre, de m'avoir fait découvrir la télédétection lors de ses cours de géomatique, de m'y avoir donné goût lors de mon mémoire, puis de m'avoir donné l'occasion de continuer dans cette voie, en m'engageant au sein de son équipe pour réaliser cette thèse. Je le remercie pour son soutien, sa confiance et ses enseignements, depuis les auditoires Croix du Sud, en passant par la Chine, pour atterrir dans les prairies de Wallonie, j'ai énormément appris. Un grand merci à Julien pour l'encadrement au quotidien. Je ne pense pas que beaucoup de doctorants puissent se vanter d'avoir travaillé dans le même bureau que leur promoteur durant toute leur thèse. Je le remercie pour sa grande disponibilité, ses nombreux enseignements, son intérêt très motivant et son implication tout au long de la thèse.

Ensuite, je remercie les membres du comité d'accompagnement et du jury, Patrick Meyfroidt, Marc Dufrêne et Benjamin Koetz, pour le temps qu'ils ont accordé à la lecture de ce manuscrit, pour leurs commentaires et suggestions, ainsi que pour les discussions enrichissantes, qui m'ont permis d'approfondir certains points et ont contribué à améliorer ce travail. Je remercie également Emmanuel Hanert d'avoir accepté de présider ce jury. Cette thèse a été réalisée dans le cadre de LifeWatch, infrastructure Européenne pour la recherche en biodiversité, et financée par la Fédération Wallonie-Bruxelles. Lifewatch aura apporté un cadre plus concret et opérationnel à cette thèse. Les réunions semestrielles avec le comité de pilotage de LifeWatch-WB, ainsi que les interactions avec nos collègues de Gembloux ont permis d'avoir des points de vue pluridisciplinaires très utiles tout au long de la thèse.

Les analyses et résultats présentés dans cette thèse n'auraient pas vu le jour sans les beaux jeux de données de terrain obtenus par différents moyens. Pour cela je remercie premièrement Arnaud Farinelle (Fourrages Mieux), pour ses mesures de rendement et d'azote, ainsi que pour sa disponibilité. Je remercie Richard Lambert et Protect'eau de m'avoir permis de prélever mes propres échantillons et le CRA-W d'avoir réalisé les mesures. Et "last but not least", je remercie Sébastien, ainsi que les collègues qui l'ont parfois remplacé, pour les longues journées de conduite à travers la campagne Wallonne, qui m'ont permis de rassembler ce super jeu de données, qui aura été la base de trois articles/chapitres et qui (à en croire Pierre) serait le meilleur en son genre

(c.-à-d les "jeux de données de terrain sur les pratiques de fauche et de pâturage en prairie").

Durant ces quatres dernières années, le groupe de vie des ENGE (R.I.P.) aura été un milieu de travail très agréable et épanouissant. Deux mots d'ordres: bienveillance et bonne ambiance. Je tiens pour cela à remercier tous mes collègues, qui ont rendu cette première expérience professionnelle exceptionnelle. J'ai été très heureuse de retrouver certains camarades d'auditoire ; Juliette, Quentin M., Lisa, Nicolas et une mention particulière à Diane pour son aide sur Sen4CAP, son énergie débordante, mais surtout ses moves incroyables. Merci aussi à Olivia pour la bonne compagnie dans le bus et pour les petites sorties course/papote en confinement; à Antoine pour la co-organisation de multiples activités et les découvertes musicales; Céline C. pour l'organisation des meilleurs events et pour ses délicieux gâteaux; à Céline B. pour les conseils voyage; à nos informagiciens Thomas et Benjamin (ou devrais-je dire St. Nicolas?) pour leur aide si précieuse; à Baptiste, super voisin de bureau, pour son enthousiasme partagé pour les burritos (n'est pas ambassadeur qui veut!); et puis Pierre H. pour le karaoke, Jean pour les anecdotes sans fin, Quentin D., Céline L., Sophie, Brigitte, Barabara, Willy, ... et les derniers arrivants: Louise, Boris, Adé... Même en se limitant aux ENGEs/géomatique, vous êtes beaucoup trop nombreux, mais merci à toutes et tous ! Entre les pauses midi, pauses café obligatoires, team building, soupers de Noël, LPS, quinzaines et autres after works, le temps passe vite ! Je tiens également à adresser mes remerciements au staff administratif, à Pascale, la maman des ENGE, et à Jasmine pour sa disponibilité et l'immense travail au sein du pôle.

Les prochains remerciements vont à toutes celles et ceux qui m'ont côtoyée, supportée, encouragée, changé les idées... durant ce chapitre de ma vie. Merci à mes amis BIR ; Oksana, Anthony, Sophie, Benoît, Marie, Florian, Régis... Merci à Aleksia pour son soutien malgré la distance et pour m'avoir fait voyager par procuration. Et un grand merci à Axelle de m'avoir challengée sportivement et maintenue en forme presque jusqu'au bout (je vais enfin pouvoir m'y remettre sérieusement). Merci aux marmottes du Refuge ; à Fif pour les heures passées à refaire le monde ; à Gillou de régulièrement me rappeler d'écouter Tool et que de toute façon la vie n'a aucun sens ; et bien sûr à Marie, toujours là après 26 ans d'amitié. Merci aux Oenopotes ; à Capu pour tous les cheese and wine ; à Clem pour ses histoires passionnantes. Et même si on se voit moins souvent, merci aux anciens de la MFR et de La Hulpe, c'est toujours un plaisir de se revoir.

Un grand merci ensuite à ma famille, sur qui je peux toujours compter dans les bons comme dans les mauvais moments. Je remercie de tout coeur mon papa et ma maman pour leur immense soutient et pour leurs nombreux encouragements. Merci à Charline d'être la petite soeur complètement folle dont j'ai besoin, et à Martin pour son humour, ses théories et philosophies, du moins celles que j'arrive à suivre. Merci aussi à Bonne Mamy, à mon parrain et à ma marraine.

Enfin, j'aimerais terminer en remerciant Dokus, qui, en entrant dans ma vie, a rendu ces dernières années de thèse tellement plus douces. Dokus, merci pour ton amour, ta présence réconfortante, ta passion contagieuse pour la nature, ton soutien au quotidien et ta patience ces derniers mois. Les perspectives apportées par nos projets de voyage m'ont boostée et motivée jusqu'au bout. J'ai hâte de pédaler jusque-là où le vent nous mène et de découvrir ce que l'avenir nous réserve.

Contents

List of acronyms and abbreviations

AECM	Agri-Environmental and Climatic Measures
API	Antecedent Precipitation Index
CAP	Common Agricultural Policy
CFAR	Constant False Alarm Rate
CIre	red edge Chlorophyll Index
CNC	Canopy nitrogen content
CNN	Convolutional Neural Network
DM	Dry Matter (yield)
DSM	Digital Surface Model
EO	Earth Observation
ES	Ecosystem Service
ESA	European Space Agency
EU	European Union
EUNIS	European Nature Information System
FAO	Food and Agriculture Organisation
fAPAR	fraction of Absorbed Photosynthetically Active Radiation
Fmask	Function of mask
GHGe	Greenhouse Gas emissions
GRD	Ground Range Detected
GUI	Grassland Use Intensity
InSAR	Interferometric SAR
IW	Interferometric Wide
LAI	Leaf Area Index
LC	Land Cover
LPIS	Land Parcel Identification System
LU	Land Use
LUI	Land Use Intensity
MAJA	MACCS ATCOR Joint Algorithm
MCC	Matthews Correlation Coefficient
ME	Mowing Event
NDRE	Normalized Difference Red Edge
NDVI	Normalized Difference Vegetation Index
NIR	Near-Infrared

NNI	Nitrogen Nutrition Index
OA	Overall Accuracy
PA	Producer Accuracy
PFA	Probability of False Alarm
REP	Red Edge Position
RMSE	Root Mean Squared Error
RTM	Radiative Transfer Model
S1	Sentinel-1
S2	Sentinel-2
SAR	Synthetic Aperture Radar
Sen4CAP	Sentinels for Common Agricultural Policy
SLC	Single Look Complex
SNAP	Sentinel Application Platform
SWIR	Short-Wave Infrared
UA	User Accuracy
VH	Polarization: Vertically sent, Horizontally received
VI	Vegetation Index
VV	Polarization: Vertically sent, Vertically received

Chapter 1

Introduction

1.1 Land use intensity: driving, mitigating and adapting to global change

Today more than 70% of the ice-free land surface has been impacted more or less intensively by human activities (Figure 1.1). Anthropogenic land use is one of the main drivers of global change (IPCC, 2021). As drivers of environmental change escalate, biodiversity declines at unprecedented rates (Butchart et al., 2010). Experts estimate that between 16 and 50% of species have been threatened or driven to extinction since the year 1500 (Isbell et al., 2022). In addition to climate change, one of the main drivers of biodiversity loss is habitat change (Duraiappah et al., 2005).

To study the extent and magnitude of those impacts, it is essential to track land cover (LC) and land use (LU) changes. LC, LU, and their evolution are crucial elements in land systems and need to be properly defined. The Food and Agriculture Organization (FAO) defines land cover as "the observed (bio)physical cover on the Earth's surface", while land use is "characterized by the arrangements, activities and inputs people undertake to produce, change or maintain a certain type of land cover" (Di Gregorio, 2005). For example, in the case of deforestation in which the LC changes from forest to grassland, from a LU perspective it

can be considered as a conversion from forestry to agriculture, or from nature conservation to food production. These are categorical changes. Continuous changes within the same class include LC modifications such as forest degradation and more subtle land use intensity (LUI) changes such as agricultural intensification.

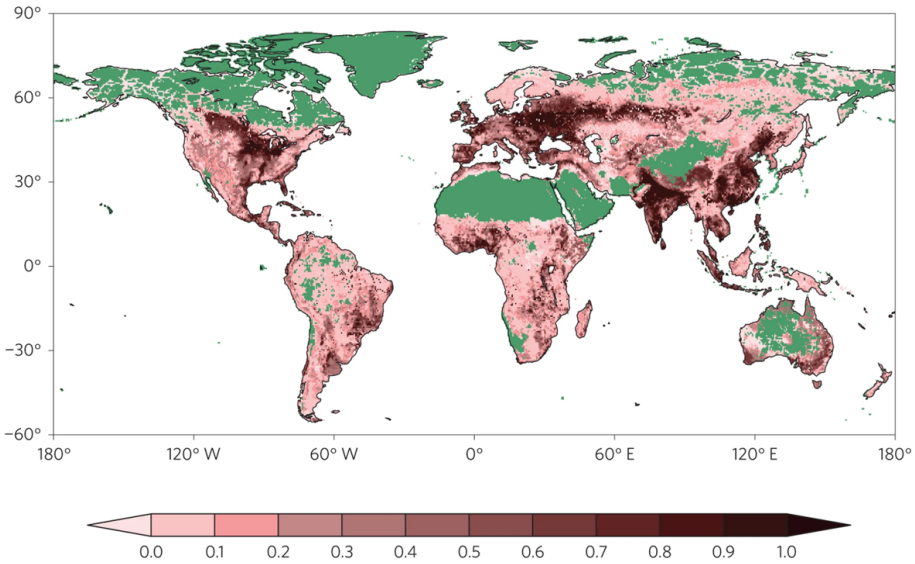

Figure 1.1: Spatial extent of land cover change, land management, wilderness and non-productive areas. Wilderness and non-productive areas are shown in green and represent land largely unaltered by humans. The remaining land is used for producing food, fiber, and fuels, and for hosting infrastructure. The colour scale represents the fraction of each grid cell for which the original plant cover was converted. Dark colours indicate regions where most of the original plant cover was converted; these regions are the subject of typical land cover change studies. The light colours show areas for which land cover change is low, but which are nevertheless under anthropogenic land management. Source: Luyssaert et al. (2014).

It has been widely shown that drastic changes in LC, such as deforestation, have impacts on climate and the Earth system in general (Brovkin et al., 2004; Ramos da Silva et al., 2008; Mahmood et al., 2014). More recently, it has become clear that more subtle LU and LUI

changes can have impacts of similar amplitudes on, e.g., surface temperatures (Luyssaert et al., 2014) or biomass and carbon stocks (Erb et al., 2018).

While the expansion of crop and pasture land (i.e. LC and LU changes) has significantly slowed down during the past years, gross agricultural production continues to increase rapidly to meet the increasing global food demand due to the growing population and per capita consumption (Rudel et al., 2009; Ramankutty et al., 2018). This increased production has been achieved mainly through agricultural intensification (i.e. increasing LUI). Agriculture is now one of the major drivers of global change. In 2019, agriculture, forestry, and other land use (AFOLU) contributed to approximately 22% of global greenhouse gas emissions (GHGe) (IPCC, 2022). About 9% of GHGe are due to land conversion (e.g. deforestation), while more or less 13% can be linked to agricultural management on already-converted lands (Ramankutty et al., 2018). Habitat fragmentation and biodiversity loss due to agricultural intensification are also major components of global change (Tilman et al., 2011). Other environmental impacts of intensive agriculture are soil degradation and water pollution.

Meeting the projected growth in global food demand without further increasing the pressure on the environment is a great challenge that will require a deep understanding of our production systems and their impacts (Tilman et al., 2011; Erb et al., 2013; Smith, 2013; Ramankutty et al., 2018; Meyfroidt et al., 2018). In this context it is crucial to consider the multidimensionality of LUI by integrating (i) input intensity, (ii) output intensity and (iii) associated system level outcome metrics (Kuemmerle et al., 2013). The conceptual framework for LUI analysis and measurement, proposed by Erb et al. (2013), systematically links those three dimensions (Figure 1.2), allowing (a) to integrate synergies and trade-offs between outputs (i.e. land-based production) and the associated unintended outcomes (e.g. biodiversity loss or GHGe), (b) to link input- and output intensity and (c) to study the relationship between changes in inputs (i.e. alternative management) and changes in system properties and their socio-ecological effects.

While our land-based production systems are a cause of global environmental change, they can also be part of the solution to mitigate those changes and become more resilient (Verburg et al., 2015). For example, a study on the impacts of LUI in grassland and forest ecosystems

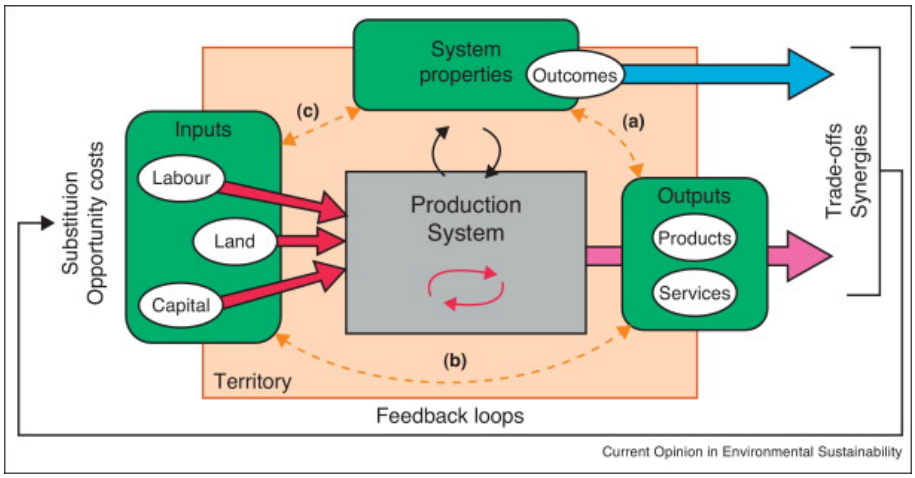

Figure 1.2: Conceptual framework for land use intensity analysis and measurement (Erb et al., 2013).

showed that positive correlations of provisioning services with biodiversity and regulating services can be strengthened up to an intermediate level of LUI, beyond which strong synergies are lost (Felipe-Lucia et al., 2020).

Furthermore, the spatial and temporal distribution of LUI cannot be overlooked. For instance, landscape structural complexity enhances biodiversity in agricultural ecosystems, which could compensate for local high-intensity management (Tscharntke et al., 2005). The structural complexity of a landscape should therefore be considered when applying agri-environmental schemes and alternative managements, as their effectiveness depends on it (Tscharntke et al., 2005; Tuck et al., 2014). Similarly, varying LUI between seasons might also be a solution to mitigate the adverse outcomes of intensification while maintaining a reasonable productivity (Allan et al., 2014).

To further understand these complex systems and work out mitigation plans, it is crucial to obtain spatially, temporally and thematically precise information on LC, LU and LUI. Over the last decades, many efforts have focused on providing global LC maps at various scales, with different levels of thematic precision (Hansen et al., 2000; Loveland et al., 2000; Arino et al., 2012; Defourny et al., 2013; Chen and Chen, 2018; Zanaga et al., 2021). There is, however, a data and knowl-

edge gap for several land management activities, such as forest harvest, nitrogen fertilization and grassland mowing and grazing (Kuemmerle et al., 2013; Erb et al., 2017). Research should therefore further focus on large-scale LU and LUI monitoring, integrating ground-based and satellite-based data to create consistent time series (Kuemmerle et al., 2013; Pongratz et al., 2018).

1.2 Grassland use intensity, ecosystem services and biodiversity

1.2.1 Grasslands of the world

Grasslands cover 20%-47% of the global ice-free land surface, depending on sources and definitions (Piipponen et al., 2022). They contribute to essential regulating ecosystem services (ES) such as carbon sequestration and water storage (Bengtsson et al., 2019; Chang et al., 2021), and they embed rich biodiversity (O'Mara, 2012; Pärtel et al., 2005; Zeller et al., 2017). Besides these essential ES, grasslands are of course a key element of most agricultural systems as they provide nearly half of the feed requirements for global livestock production, and account for 69% of the world's agricultural area (O'Mara, 2012; Herrero et al., 2013).

Grasslands encompass a wide range of LC classes and occur in various ecoregions, spread across all continents (except Antarctica) (Dixon et al., 2014). Depending on scopes and perspectives, there exist numerous definitions for grasslands. In the FAO land cover classification system (LCCS), grasslands in a broad sense are defined as *"A12/A2: natural and semi-natural herbaceous vegetation"*. The term "herbaceous" designates non-woody vegetation, including forbs, graminoids, lichens and mosses (Di Gregorio, 2005). This very inclusive definition is often used as a basis for global grassland definitions. For instance, Dixon et al. (2014) defined grasslands as *"a non-wetland type with at least 10% vegetation cover, dominated or co-dominated by graminoid and forb growth forms, and where the trees form a single-layer canopy with either less than 10% cover and 5 m height (temperate) or less than 40% cover and 8 m height (tropical)."* These global definitions of grasslands include a variety of LC classes, ranging from very sparsely vegetated semi-arid grasslands, to temperate grasslands, and tropical savannas with up to 40%

canopy cover (trees or shrubs). Dixon et al. (2014) identified 49 taxonomically and spatially distinct grassland types, including i.a. steppes, savannas, and alpine meadows.

Grasslands are also diverse in terms of origin, LU, and LUI. Commonly, based on their origin and LU, a distinction is made between natural and semi-natural — or secondary — grasslands (Di Gregorio, 2005; Dengler et al., 2020). Natural grasslands occur naturally due to pedo-climatic conditions, are balanced, and mostly undisturbed (Di Gregorio, 2005). According to some definitions, the current state of these grasslands can however be modified by human LU (Dengler et al., 2020). Secondary grasslands, on the contrary, occur in places where the natural vegetation is e.g. forest or wetland, and originated and are maintained, through human LU like mowing, grazing, burning, or abandoning arable fields (Dengler et al., 2020). Some more intensive grasslands are even seeded and included in crop rotations (Peeters et al., 2014). The latter (i.e. temporary grasslands and fodder crops) are however often considered as cropland rather than grassland, especially in definitions with a global scope (Di Gregorio, 2005; Dengler et al., 2020). Undisturbed natural grasslands are mostly grazed by wild ungulates. The most common anthropogenic use of grasslands worldwide is livestock grazing. The production systems and associated LUI range from extensive nomadic pastoral activities to strongly intensified, sown and fertilized grassland parcels with high stocking densities or mowing frequencies.

1.2.2 Grassland ecosystem services: synergies and trade-offs

Grasslands deliver crucial provisioning and regulating ES, including food production, wildlife habitat, carbon sequestration, and water storage. Over the past decades, grasslands have been used more and more intensively to meet the increasing demands of dairy and meat production, often at the expense of climate and biodiversity. Adequate management can, however, create synergies and balance trade-offs between provisioning and regulating ES.

Globally, grasslands play an often underestimated role in carbon storage (Bai and Cotrufo, 2022), as they store approximately 34% of the terrestrial carbon stock, mostly below ground, as soil organic carbon and root biomass (White et al., 2000; O'Mara, 2012). Chang

et al. (2021), however, showed that the global warming caused by the GHGe of intensively managed grasslands cancels the climate cooling from carbon sequestration in natural and semi-natural grasslands. The main drivers of intensive grassland GHGe are manure and fertilization management and CO_2 and CH_4 emissions from livestock production (Chang et al., 2021). The livestock sector in general is responsible for a significant part of global GHGe and represents more than half of the technical mitigation potential of the AFOLU sectors (Herrero et al., 2016). Grasslands carry a significant part of this mitigation potential (Bai and Cotrufo, 2022). Adequate grassland management through — e.g. conversion from cultivation to grasslands, increasing plant diversity, sowing legumes and grasses, fertilization, and adapted grazing rate to maximize forage production — can enhance soil carbon storage, reducing the net GHGe of grasslands (Herrero et al., 2016; Chang et al., 2021; Bai and Cotrufo, 2022).

Grasslands also play an essential role as habitat for a host of plant and animal species and thereby present a great potential for the conservation of biodiversity. Grasslands ES and biodiversity are however strongly influenced by their LU and LUI (Hudewenz et al., 2012; Allan et al., 2014; Chisté et al., 2016; Van Vooren et al., 2018). There is a large ecological value gradient between species-rich natural grasslands and intensively used mono-specific temporary grasslands and fodder crops. Worldwide, natural grasslands are threatened by climate change and conversion to arable land (Dengler et al., 2014). In agricultural landscapes, semi-natural grasslands of high ecological value are threatened by intensification or abandonment with potential woody plant encroachment.

Many studies have related grassland biodiversity to their LUI (Hudewenz et al., 2012; Allan et al., 2014; Dengler et al., 2014; Chisté et al., 2016; Van Vooren et al., 2018). For example, an exponentially declining relationship was found between nitrogen input and grassland plant species richness (Kleijn D et al., 2009). Similarly, the timing and frequency of mowing events can have an impact on certain bird species' abundance. In agricultural landscapes, heterogeneous mowing dynamics optimize and maintain flower resources for pollinators along the season (Horn and Koford, 2000; Johansen et al., 2019).

However, studies conducting meta-analyzes on the effects of LUI on biodiversity found very heterogeneous results, concluding that the

links between LUI and grassland species richness are habitat-specific. There also appears to be no clear consensus on the difference in impact between grazing and mowing, in terms of species richness. While many consider extensive grazing as a good practice, as the grazing, trampling and defecation patterns create small-scale heterogeneity, others showed that extensive mowing regimes were more effective to enhance biodiversity (Dengler et al., 2014). Pedo-climatic conditions, site-specific potential productivity, and landscape elements need to be considered for the determination of sustainable management practices (Humbert et al., 2012; Tälle et al., 2018). For example, Shahan et al. (2017) found that the occurrence of grassland songbirds on remnant grassland patches was more linked to landscape characteristics than to local management practices. Furthermore, varying LUI across years can also enhance grassland biodiversity (Allan et al., 2014; Johansen et al., 2019).

Even in a given agro-pedo-climatic context, it is unlikely that there is a single management solution that optimizes all ES since some trade-offs are inevitable (Bengtsson et al., 2019; Savage et al., 2021). However, studies showed that intermediate intensity levels can create synergies and limit trade-offs among ES (Felipe-Lucia et al., 2020), establishing floristically diverse grasslands enhances multiple ES, and appropriate cutting management can boost resources for birds and natural predators of pests without deteriorating yield and other ES (Savage et al., 2021).

1.2.3 Grassland use intensity: definitions and measurement

The diversity of grasslands worldwide, and the discrepancies between definitions and typologies depending on the scope and perspectives, can lead to confusion and misinterpretations between studies. Before defining grassland use intensity (GUI) in the framework of this thesis, it is, therefore, necessary to narrow down the scope, and clearly define which types of grasslands are considered.

The focus is set on managed grasslands with a closed vegetation cover and less than 10% woody plant cover. In this context, the term "managed" refers to sedentary anthropogenic activities with agricultural or ecological purposes. For the sake of readability, unless specified otherwise, the term "grassland" will refer to this definition in the

following text of the manuscript. This includes e.g. natural grassland areas, such as alpine pastures, that are grazed for livestock production, semi-natural grasslands that are grazed and/or mown for biodiversity conservation, as well as highly productive permanent grasslands used as hay meadows or pasture, and intensively fertilized temporary grasslands. On the contrary, natural grasslands with no activity, sparse semi-arid grasslands, and savannas with nomadic pastoral activities are not considered here. Geographically, the focus is primarily set on Europe, as this type of grassland are dominant in Western, Eastern, as well as Mediterranean parts of Europe (Dengler et al., 2020). However, this type of grassland can also be found in other regions and climates.

Despite the narrower scope, this grassland definition still encompasses a large diversity, both from an ecological and agricultural point of view. The state of grasslands and their ES largely vary depending on pedo-climatic conditions and GUI. In the literature, most studies on GUI focus on management practices. The type and quantity of management practices, particularly grazing, mowing and fertilization, are commonly used as indicators to classify grasslands, both from an agricultural and an ecological perspective (Blüthgen et al., 2012; Peeters et al., 2014; Tonn et al., 2020).

For example, the GUI indicator of Blüthgen et al. (2012) is calculated by summing up three regionally standardized GUI factors, namely mowing (in cuts/year), grazing (in livestock units grazing days/ha.year), and fertilization (in kg N/ha·yr). This index has been widely used to measure GUI (Allan et al., 2014; Vályi et al., 2015; Chisté et al., 2016; Van Vooren et al., 2018; Busch et al., 2019). However, some studies found it was insufficient to study the impacts of management practices on ecosystem services (ES) and biodiversity, and it could be improved by adding more variables (e.g. first mowing date) and assigning weights to the different components (Van Vooren et al., 2018).

More recently, Tonn et al. (2020) established a management-based typology for European permanent grasslands, i.e. grasslands that have not been included in an arable crop rotation for at least five years (Figure 1.3). In this typology, managed permanent grasslands with less than 10% woody plant cover and a renewal interval of more than 15 years are classified by management intensity level, using the nitrogen fertilizer input, the number of cuts per year, and the stocking rate of grazing animals as indicators. To integrate site-specific productiv-

ity potential, which has an impact on the outputs and outcomes of a given management intensity level (Erb et al., 2013), two classes of pedo-climatic productivity potential are considered (productive- and marginal region). Finally, grasslands are further differentiated based on the dominant exploitation practice, namely grazing or mowing.

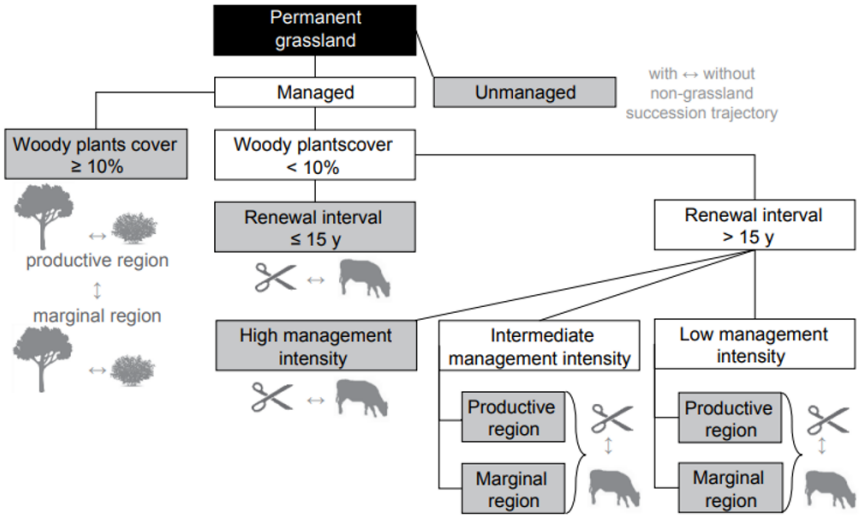

Figure 1.3: Schematic overview of a management-based typology for European grasslands. First-level classes of the typology are in grey boxes, subordinate second-level classes are indicated either by text or by two different sets of symbols (tree- vs shrub-dominated; predominantly cut vs predominantly grazed) (Tonn et al., 2020).

In previous examples, GUI measurement is limited for the most part to three management practices: mowing, grazing, and fertilization. Other practices, such as irrigation, ploughing, and seeding could be considered in GUI measurement as well. Moreover, the impacts of management practices are site-specific and depend on the agro-pedo-climatic context (Grigulis et al., 2013; Abdalla et al., 2018; Klein et al., 2020). Therefore, characterizing and quantifying management practices alone is not sufficient. For example, the stocking density of grazing livestock should be balanced against the site-specific carrying capacity, i.e. the number of grazing animals a piece of land can support (Piipponen et al., 2022). Based on the LUI analysis framework of Erb

et al. (2013) (Figure 1.2), GUI should also be declined in three dimensions. The input intensity is then defined by the type and quantity of management practices (mowing, grazing, fertilization, ploughing, irrigation ...), the output intensity could be measured through grassland productivity (e.g. in terms of forage yield and quality) and finally, the impacts on biodiversity and regulating ES are system-level outcomes.

1.2.4 Grassland mapping and monitoring

The ecological state and condition of managed natural and semi-natural grasslands are threatened by agricultural intensification (O'Mara, 2012; Silva et al., 2008). Appropriate and spatially optimized management practices integrating knowledge of ecological processes are key for creating synergies and balancing trade-offs among the food production on one hand and regulating ecosystem services and biodiversity of grasslands on the other (Bengtsson et al., 2019; Chang et al., 2021; Pärtel et al., 2005; Savage et al., 2021). It is therefore of great interest to measure, map and monitor GUI to a large extent and with a sufficient spatial and temporal resolution to study the effects of changes in GUI and guide the design of adequate agricultural policies.

In Europe, several efforts aim at collecting LU, LUI, and more specifically GUI data at different levels of spatial, temporal, and thematic precision and coverage.

In the frame of the European common agricultural policy (CAP), all Member States use the land parcel identification system (LPIS) to record all agricultural parcels (Owen et al., 2016). The LPIS is provided as a vector dataset based on CAP declarations by farmers in each EU country, including parcel boundaries and crop types. Depending on the country or region, the thematic precision of the LPIS varies greatly. While some include one single grassland class, others provide information on the type (temporary or permanent), the use (e.g. pasture or hay meadow), and even the use intensity or the biodiversity value of grasslands. Maintaining permanent grasslands is one of the main concerns of the European CAP.

In the framework of the new CAP of 2023-2027, grasslands and their management play an essential role in multiple key objectives, including the contribution to climate change mitigation, halting and

reversing biodiversity loss, and more specifically maintaining and enhancing pollinator populations and diversity. The preservation of permanent grasslands is a major aim of the CAP baseline standards, which set a threshold for conversion to other agricultural uses and prohibit conversion and ploughing in "environmentally-sensitive permanent grasslands" in Natura 2000 sites. Furthermore, the agri-environmental and climatic measures (AECM), i.e. subsidy schemes to encourage sustainable management practices, include the maintenance of semi-natural and high nature-value grasslands.

From an ecological perspective, European, national and regional initiatives carry out field surveys focussed on habitat characterization and monitoring (e.g. Dufrêne and Delescaille (2005)). The European nature information system (EUNIS) is often used as basis for habitat characterization. It describes the *"Grasslands and lands dominated by forbs, mosses or lichens (E)"* habitat as *"Non-coastal land which is dry or only seasonally wet (with the water table at or above ground level for less than half of the year) with greater than 30% vegetation cover. The vegetation is dominated by grasses and other non-woody plants, including mosses, macrolichens, ferns, sedges and herbs. Includes semiarid steppes with scattered Artemisia scrub. Includes successional weedy vegetation. Excludes regularly tilled habitats dominated by cultivated herbaceous vegetation such as arable fields and artificial grasslands and herb dominated habitats."* On a second level the EUNIS differentiates seven types of habitats, including "dry grasslands (E1)", "seasonally wet and wet grasslands (E3)" and "mesic grasslands (E2)". Mesic grasslands further include "mesic permanent pasture of lowlands and mountains (E2.1)" and "low and medium altitude hay meadow (E2.2)" which are the most widespread traditionally managed grasslands in temperate Europe. These habitats have been widely transformed by agricultural intensification over the past decades (Chytrỳ et al., 2020).

The Land Use/Cover Area-Frame Survey (LUCAS) is a European field survey providing information on land use, land cover, and environmental parameters every three years on a statistically representative sample of points spread across the EU countries (d'Andrimont et al., 2020). Recently, a new module was added to the survey, specifically focusing on grasslands (Sutcliffe et al., 2019). It aims to characterize habitat types (EUNIS), environmental conditions, age of grasslands, use type, use intensity, vegetation structure, biodiversity value, and

pollinator value. During the pilot of this new module in 2018, 3734 grassland points in 26 countries were recorded on the field by trained surveyors following a standard methodology.

These existing datasets are often limited temporally, geographically, or thematically. Ground-based grassland monitoring methods, such as visual assessment, floristic inventories, clipping, and field spectrometry are very useful at the local scale, but they are much less feasible over large areas or at regular intervals since they are either subjective or very costly and time consuming (Ali et al., 2016a). Remote sensing can therefore be a great asset. The increasing spatial resolution and revisit frequency of available satellite time series presents a considerable potential for large-scale grassland monitoring (Ali et al., 2016a; Reinermann et al., 2020). High-quality field measurements however remain essential for calibrating and validating satellite-based mapping and monitoring methods.

1.3 Remote sensing for grassland monitoring

Remote sensing has allowed tremendous progress in the field of LC, LU, and LUI mapping and monitoring. Especially over the past decade, satellites such as the Copernicus Sentinel missions provide extremely valuable data with global coverage and increasing spatial and temporal resolution. By measuring reflected (optical sensors), emitted (thermal sensors) or backscattered energy (active radar sensors), remote sensors record different electromagnetic properties of the Earth's surface, which can be linked to structural, chemical, or biophysical properties (Joshi et al., 2016).

1.3.1 Optical and microwave sensors

Optical remote sensing is largely and successfully used in agricultural mapping and monitoring (Ali et al., 2016a; Reinermann et al., 2020). The different wavelengths of multi-spectral optical sensors (commonly ranging from the visible to the near-infrared (0.38 - 2.5 μm)) allow to reconstruct the spectral signature of a vegetation cover. Figure 1.4 shows a typical grassland reflectance spectrum.

Different parts of the spectrum can be linked to specific properties of a vegetation cover, and vegetation indices (VI) combining mul-

tiple wavelengths are used to accentuate particular spectral features for the retrieval of specific properties. The visible part of the spectrum (0.45 - 0.66 μm) is commonly linked to pigments, such as chlorophyll and carotenoids (Hank et al., 2019). The red (0.62 - 0.70 μm) and near-infrared (NIR, 0.75 - 1.3 μm) reflectances are strongly related to biophysical quantities, such as e.g. the biomass and the leaf area index (LAI). These reflectances are used to compute the very popular Normalized Difference Vegetation Index (NDVI) (Rouse Jr et al., 1973), which is closely linked to green biomass. The red edge (i.e. the rapid increase in reflectance from the red to NIR (around 0.68 - 0.78 μm)) is particularly sensitive to vegetation chlorophyll content (Filella and Penuelas, 1994; Clevers and Gitelson, 2013; Hank et al., 2019). Finally, the short-wave infrared (SWIR) domain (around 0.9 - 2.5 μm) is mostly influenced by water content and lignin, cellulose, and senescent material (Hank et al., 2019).

More than the reflectance spectrum, the evolution of spectral features and indices over time is key in agricultural monitoring. Temporal profiles extracted from image time series can be used to classify crop types (Belgiu and Csillik, 2018; Defourny et al., 2019) and monitor agricultural practices (Ottosen et al., 2019).

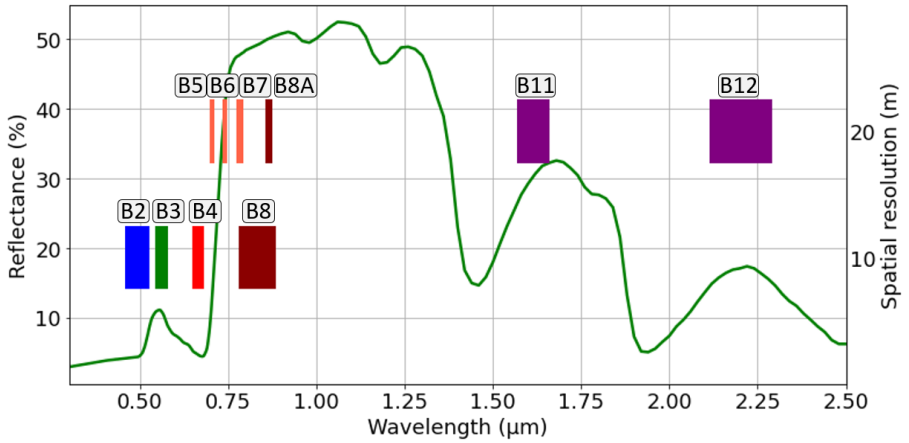

Figure 1.4: Sentinel-2 10 m and 20 m resolution band settings compared to a typical grassland reflectance spectrum (Johns Hopkins University, 2018).

The launch of the Sentinel-2 (S2) A and B satellites has offered unprecedented capacities in this domain, as it combines a high spatial resolution (10 to 20 m) with a short global revisit time (10 days each, 5 days together) and a fine spectral resolution (Figure 1.4). The multispectral imager of S2 measures surface reflectance in blue (B2), green (B3), red (B4), and NIR (B8) with a 10 m resolution. Six other bands are measured with 20 m resolution, including three vegetation red edge bands (B5-7), a narrow NIR band (B8A), and two SWIR bands (B11-12). Three additional spectral bands (B1, B9, and B10) are intended for atmospheric corrections with a spatial resolution of 60 m.

Despite the short revisit time of recent satellites, obtaining sufficiently regular and frequent observations can be a challenge when using optical sensors in regions with frequent cloud cover (Kolecka et al., 2018; Sano et al., 2007). Synthetic aperture radar (SAR) satellites carry active sensors, sending electromagnetic microwaves to the Earth's surface and measuring the backscattered signal amplitude and phase in different polarizations. Since active sensors are independent of sunlight and their microwaves can pass through cloud covers, SAR satellites provide regular observations through day and night. The backscattered signal amplitude depends on sensor parameters (wavelength, polarization, and incidence angle) and ground parameters (geometry, orientation, and dielectric constant of soil and objects on the surface). A given signal interacts with objects and surface roughness with sizes larger or equal to its wavelength. C-band radars, such as Sentinel-1 (S1), emit a signal with a frequency of 5.405 GHz, corresponding to a wavelength of about 5.55 cm. They penetrate vegetation covers and interact with multiple structural elements (leaves, branches), before being scattered back to the sensor. This backscattering mechanism is referred to as volume scattering. On bare soil or short vegetation, the backscattered signal amplitude is influenced mostly by surface scattering, as the signal interacts with fewer elements. Depending on the roughness and the composition of the surface, the scattering is more or less diffuse. On a smooth surface, such as still water, the incoming signal is almost completely scattered back in another direction (i.e. specular scattering), so only a very weak signal is received by the sensor. On surfaces with roughness around the size of the radar wavelength, the incoming signal will be scattered in multiple directions. Depending on the orientation of scatterers, a larger part of the signal is scattered back

to the sensor. These scattering mechanisms are schematized in Figure 1.5.

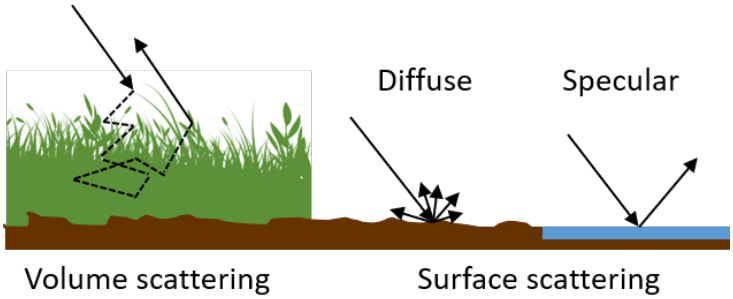

Figure 1.5: Schematic representation of SAR scattering mechanisms on different surfaces.

The electromagnetic signal emitted by SAR antenna is polarized, often horizontally (H), or vertically (V). Depending on the target's shape, roughness, orientation, and dielectric properties, the backscattered signal either remains polarized in the same direction or is re-polarized in the opposite way. The sensors also measure the backscattered signal in different polarizations (H or V). The degree of re-polarization can thereby be estimated and provide additional information on the target's properties.

Interferometric SAR (InSAR) measures the phase difference between two radar observations of the same area, taken from slightly different look angles. The interferometric coherence, which is a cross-correlation coefficient of two consecutive SAR observations, was initially computed to estimate phase noise for interferogram quality assessment. It has, however, also been exploited directly to estimate the temporal stability of ground targets. InSAR coherence has thereby been used for various applications including land cover mapping (Jacob et al., 2020; Strozzi et al., 2000), crop monitoring (Blaes et al., 1999; Shang et al., 2020) and soil moisture estimation (Barrett et al., 2009; De Zan et al., 2013; Rabus et al., 2010; Ulaby et al., 1979).

S1 is the first satellite constellation of the Copernicus Programme (conducted by the European Space Agency). S1 A and B carry C-band SAR sensors. The interferometric wide swath acquisition mode offers a spatial resolution of 5 × 20 m in dual polarization (VV + VH). In VV or

co-polarization, the sensor receives polarized waves in the same polarization as it sent them out (Vertically sent – Vertically received). In VH or cross-polarization, it receives polarized waves in the opposite polarization as it sent them out (Vertically sent – Horizontally received). The cross-polarization backscattering is particularly useful for vegetation monitoring, as volume scattering tends to re-polarize the signal significantly.

1.3.2 Remote sensing of grasslands: overview

As for agricultural monitoring in general, time series with sufficient temporal resolution are crucial for grassland monitoring, to capture the different phenological stages and management practices. Until recently, short revisit times were associated with lower spatial resolutions, preventing field-scale applications (Ali et al., 2016a). In recent years, with the emergence of new satellites combining high spatial and temporal resolution, such as the Sentinel missions, an increasing number of studies have shown the potential of remote sensing for grassland mapping and monitoring (Ali et al., 2016a; Reinermann et al., 2020). Figure 1.6 provides an overview, illustrating the major aspects of grassland monitoring through remote sensing.

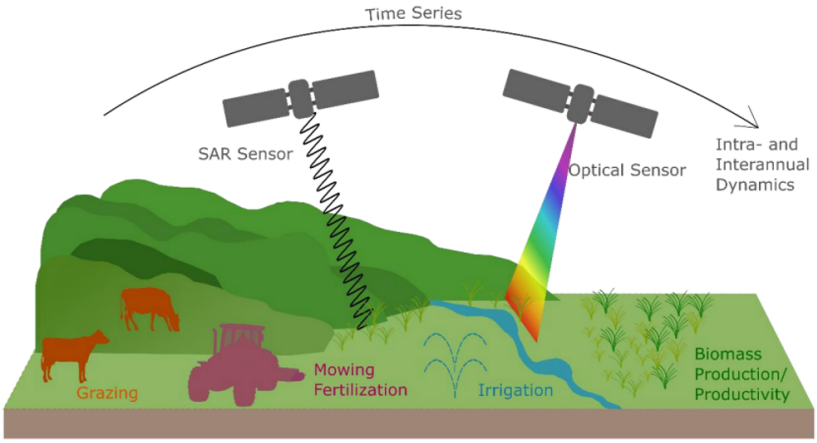

Figure 1.6: Overview of satellite remote sensing (optical and SAR) of major drivers and processes in managed grassland ecosystems (Reinermann et al., 2020).

Most studies on grassland monitoring focus on outputs of GUI, through the retrieval of biomass or associated biophysical variables (e.g. LAI), from reflectances and derived indices (Sibanda et al., 2015; Zhang et al., 2015; Ali et al., 2016b; Quan et al., 2017; Reinermann et al., 2020; Hardy et al., 2021). Remote sensing image time series are, however, increasingly used to map management types and intensities (Stumpf et al., 2020). Grassland management monitoring has been explored, either through image classification methods (Franke et al., 2012; Barrett et al., 2014; Dusseux et al., 2014) or by retrieving different factors of GUI (i.e. mowing, grazing or fertilization) separately from image time series and auxiliary data (Asam et al., 2015; Gómez Giménez et al., 2017; Estel et al., 2018; Stumpf et al., 2020; Hardy et al., 2021). Finally, some studies discriminate grassland habitats, based on remote sensing image classification (Rapinel et al., 2019; Fazzini et al., 2021; Kaasiku et al., 2021).

Overall, recent studies on grassland monitoring show promising results. However, they are often conducted on relatively small study areas or lack precise and complete reference data for representative performance evaluation (Ali et al., 2016a; Reinermann et al., 2020).

1.3.3 Grassland use intensity measurement methods

Many studies focus on the retrieval of grassland biophysical variables, most of which estimate biomass and yield (Sibanda et al., 2015; Zhang et al., 2015; Ali et al., 2016b; Quan et al., 2017; Hardy et al., 2021). Biomass and other variables can be related to spectral bands and indices using linear or more complex regression models (Sibanda et al., 2015; Dusseux et al., 2015; Ullah et al., 2012; Wang et al., 2019). Ullah et al. (2012) used MERIS data to retrieve grassland dry biomass and nitrogen content through a stepwise multiple linear regression approach with respective Root Mean Squared Error (RMSE) of 1.36 t/ha and 42 kg/ha. Radiative transfer models (RTM) are increasingly used to retrieve biomass and other biophysical variables (Quan et al., 2017; Schwieder et al., 2020). Schwieder et al. (2020) showed that both empirical regression and RTM approaches are suitable for grassland aboveground biomass estimation. The normalized RMSE on their small study site was 47% with the RTM and 17% with the empirical model.

Fewer studies have tested the retrieval of nitrogen content (Adjorlolo et al., 2014). Adjorlolo et al. (2014) retrieved grassland N content from multispectral WorldView-2 data using random forest regression, reaching an r^2 of 0.68 with the NIR and red edge bands. Pullanagari et al. (2021) used field spectroscopy and convolutional neural networks to retrieve canopy nitrogen contents on 17 temperate grasslands with a normalized RMSE of 14%. Their analysis showed that the most determinant bands were in the red edge, NIR, and SWIR part of the spectrum.

Grassland management practices or input intensity, have been assessed and monitored by remote sensing through different kinds of approaches. Franke et al. (2012) and Asam et al. (2015) used 5 m resolution RapidEye image time series to directly classify grasslands into a predefined number of management intensity classes, ranging from semi-natural to intensively managed and even tilled grasslands and from pastures to hay meadows. While Asam et al. (2015) lacked sufficient field data for a quantitative evaluation, Franke et al. (2012) estimated an overall accuracy of 85% for their decision tree classification using 1500 field data points drawn from a small study area (500 km^2) in Germany. The classification was based on per parcel NDVI and Normalized Red-Edge Vegetation Index time series derived from 5 RapidEye images. Both studies highlight the importance of dense time series to estimate GUI.

Different aspects of grasslands input intensity can also be measured as independent variables. Mowing frequency, grazing intensity, and fertilization input have been estimated using high-resolution multispectral image time series (MODIS, RapidEye, S2) and agricultural statistics, regionally (Gómez Giménez et al., 2017) or at a broader scale (Estel et al., 2018). Gómez Giménez et al. (2017) estimated mowing dates based on change detection in red edge vegetation index ($VI_{rededge}$) time series extracted from 5 RapidEye images. The coefficient of variation (CV) of the $VI_{rededge}$ time series was used to retrieve grazed parcels and livestock density was used as a proxy for fertilization input. They could not validate their method, by lack of field data, but obtained coherent results in terms of mowing and grazing activities. No correlation was however found between livestock density and spectral response. Estel et al. (2018) estimated GUI at the European scale, using MODIS (250 m) NDVI time series to detect mowing events at the landscape

level and regional livestock statistics. Their results were regionally coherent but not quantitatively validated.

Recently, in the field of GUI monitoring, particular attention has been given to mowing detection. A mowing event implies a drastic change as a significant fraction of biomass is removed from the surface. Such a change has an impact on the absorption, reflection, and scattering of incoming radiation by vegetation. It is thereby theoretically possible to monitor mowing events through change detection using optical or SAR image time series with sufficient spatial and temporal resolutions. In optical remote sensing, mowing events have been associated with sudden decreases in NDVI profiles (Estel et al., 2018; Kolecka et al., 2018; Griffiths et al., 2020; Schwieder et al., 2022) or other vegetation indices related to biomass (Gómez Giménez et al., 2017).

For short-term change detection it is crucial to ensure a regular and dense acquisition frequency, which can be a challenge when using optical sensors in cloudy regions (Kolecka et al., 2018; Griffiths et al., 2020). Therefore, it is interesting to assess the potential of radar remote sensing for mowing detection. SAR backscattering is influenced by surface structure and humidity. Cutting grass changes the vegetation structure and should thereby influence the radar signal backscattering. Schuster et al. (2011) established a mowing detection rule for a NATURA-2000 habitat monitoring and surveillance scheme, based on two axioms, namely the occurrence of a backscattering signal rise, followed by a signal decrease and a threshold magnitude of the signal changes. The use of Artificial Neural Networks on S1 backscattering data for detecting mowing events was also explored, showing promising results on 5 parcels in Germany (Taravat et al., 2019).

In addition to backscattering coefficients, the interferometric coherence between two SAR acquisitions can also be used for mowing detection (Tamm et al., 2016; Zalite et al., 2016). On tall vegetation, the signal is dominated by volume scattering. Due to the gradual growth of vegetation and random movements of tall grass in the wind, the distribution of scatterers changes between two acquisitions, causing temporal decorrelation of the signal and hence lower coherence values (Blaes et al., 1999; Monti-Guarnieri et al., 2020; Morishita and Hanssen, 2014; Voormansik et al., 2020). Zalite et al. (2016) found an inverse logarithmic relation between the temporal interferometric coherence and the vegetation height and wet above-ground biomass. InSAR coherence is

relatively low on grasslands during their growing phase. After a mowing event, as the grass is cut short, the soil surface scattering, which is more stable over time, dominates the signal (Blaes et al., 1999). The coherence is therefore higher. Typical coherence profiles, showing lower values during the growing phase and high values after the mowing event, have been observed on aggregated time series of mown grasslands (Tamm et al., 2016).

1.3.4 Knowledge gaps and short-comings

Remote sensing offers great potential for large-scale and cost-effective grassland monitoring. The unprecedented spatial and temporal resolution of Sentinel-1 and Sentinel-2 have opened up new possibilities in this field. Recent studies have already shown the feasibility of measuring different aspects of GUI, through various approaches and with different levels of precision. Further research is, however, needed to bridge knowledge gaps on the relationships between optical or microwave signal and grassland properties, to improve measurement methods, to assess their potential and identify their limitations, and to move from LC and LU mapping toward operational large-scale grassland use intensity monitoring.

Most grassland monitoring studies, although showing promising results, were performed on relatively small study areas, or lacked sufficient reference data for a robust performance assessment. Studies on grassland biophysical variable retrieval, for example, were mostly based on empirical models which are locally tuned or calibrated on small datasets covering short periods. Grassland management monitoring, such as mowing detection methods, were also mostly calibrated and validated on small reference datasets, or could not be validated quantitatively at all. Moreover, many studies focus on one type of grassland and are not representative of the diversity of grassland types and management practices. Most studies on grassland mowing detection for example focus only on mown meadows, not taking pastures and mixed practices into account.

The most frequently used index for GUI measurement seems to be the NDVI, followed by biophysical indices such as the LAI. In general, optical remote sensing is used in more than 90% of research on GUI

measurement (Reinermann et al., 2020). Overall, there is a better understanding of the relations between reflectance and vegetation properties, compared to radar remote sensing. Microwave data, such as S1 C-band, can nevertheless be used instead of - or in addition to - optical imagery since they guarantee regular temporal coverage and provide complementary information. The latest studies on mowing detection with SAR imagery provide encouraging results, but have mainly been carried out on rather limited study areas, covering between 2 and 40 grassland parcels (Schuster et al., 2011; Tamm et al., 2016). Moreover, while radar acquisitions are completely independent of cloud coverage, it is preferable to take precipitation into account, as the signal backscattering is largely influenced by surface humidity and rain on the acquisition date might interfere with the actual vegetation monitoring (Curnel, 2015; Tamm et al., 2016).

The choice of a spatial unit represents another issue. One of the main challenges in SAR is the speckle, an inherent variance caused by constructive and destructive interference between randomly distributed scatterers within a pixel (Lee et al., 1994). Homogeneous surfaces such as intensively managed herbaceous covers, therefore, appear heterogeneous, which makes it challenging to work at the sub-parcel level. Therefore, it is useful to smooth the signal by averaging the remotely sensed metrics per parcel, which is the case in most studies on GUI measurement with SAR. However, when working by object, potential intra-parcel heterogeneity in management practices becomes an issue. Further research should investigate the relationship between SAR signal and grassland properties and the above-mentioned challenges need to be tackled to develop robust SAR-based grassland monitoring methods.

1.4 Scope and objectives

A better characterization of grassland production systems is essential to evolve towards more sustainable grassland management, i.e. adapting inputs to create synergies and balance trade-offs between outputs and outcomes of grassland ecosystems. Therefore, temporally and spatially explicit data on each aspect of GUI are crucial. Existing datasets are however often limited temporally, thematically, or geographically and collecting data on the field is extremely costly and time-consuming.

The overarching objective of this thesis is to measure grassland use intensity over large areas thanks to satellite remote sensing. Based on a general LUI measurement framework (Erb et al., 2013) and considering existing GUI indices and typologies (Blüthgen et al., 2012; Tonn et al., 2020), we propose a new framework for GUI measurement (Figure 1.7). GUI is articulated in three dimensions, namely inputs (mowing, grazing, and fertilization), outputs (forage yield and quality), and outcomes (impact on biodiversity and regulating ES). This framework is primarily intended for managed grasslands, as defined in section 1.2.3, and focuses on the most common management practices in Europe.

Although they are all important to consider for GUI analysis, not all aspects shown in Figure 1.7 are directly addressed in this thesis. The focus is set on inputs, more specifically on management practices (i.e. mowing and grazing), and on outputs through forage yield retrieval. The goal is to provide accurate and precise information on these aspects, which can allow to derive other aspects, relate GUI inputs and outputs with outcomes, and contribute to large-scale agricultural and ecological monitoring.

Previous studies on GUI measurement by remote sensing have mostly been conducted on relatively small areas and rarely cover the diversity of grasslands. One of the main limiting factors in this field is the validation, as reference datasets are often missing, not representative, or lack spatial and temporal resolution. Empirical methods are therefore often locally tuned and findings can not be generalized. Further efforts are needed to provide a comprehensive and quantitative evaluation of GUI measurement methods and improve their performances. In this thesis large and comprehensive field datasets are used to develop, validate and combine change detection-, classification- and retrieval

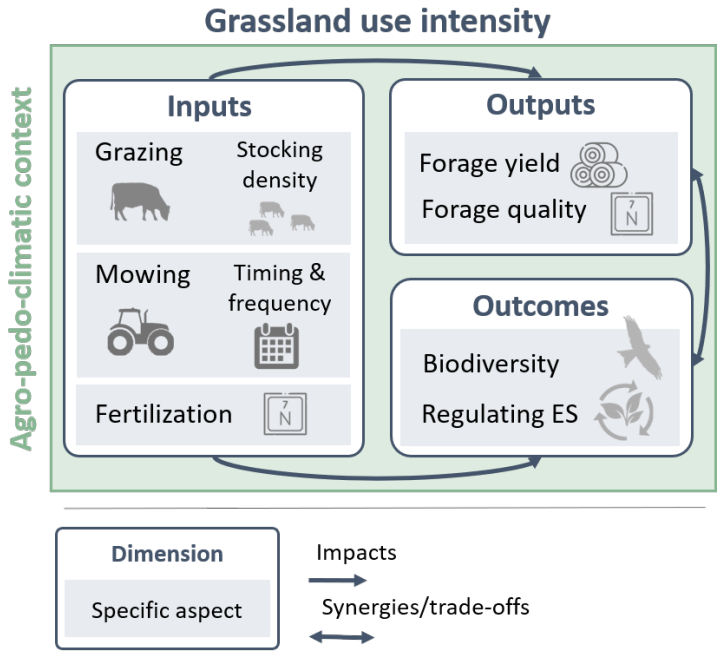

Figure 1.7: Conceptual framework for grassland use intensity analysis and measurement considered in this research.

methods based on microwave and optical remote sensing time series and existing ancillary data.

The first specific objective is to differentiate grassland mowing dynamics by retrieving the timing and frequency of mowing events. Mechanical mowing plays a major role in European grassland management from a food production and habitat conservation perspective. Mowing events cause an abrupt reduction in above-ground biomass, followed by gradual regrowth. Sufficient temporal resolution is key to detect such punctual changes. Therefore, an initial mowing detection method is developed based on microwave remote sensing. In this first study, the potential of S1 for mowing detection is assessed, and major confounding factors are identified. Secondly, an automated mowing detection method combining the completeness of S1 and the higher accuracy of S2 time series is presented and evaluated. In this second study, the complementarity of S1 and S2 is assessed and the performances of the combined mowing detection method are evaluated

across regions and grassland types to test its potential for large-scale grassland monitoring.

The second specific objective is to differentiate grazing and mowing practices at the sub-parcel level, corresponding to homogeneous grassland management units. Since we focus on managed grasslands, we assume that all parcels are grazed and/or mown at least once a year. Most European temperate grasslands are managed by mowing, grazing, or a combination of both within declared parcels of the LPIS. In the frame of GUI assessment, grazing and mowing can however not be considered equivalent. While a mowing event causes an abrupt decrease in biomass, grazing events are often more gradual (except in intensive rotational grazing systems). Moreover, in most pastures, selective grazing, trampling, and defecation patterns create small-scale heterogeneity, whereas mechanical mowing is more uniform and unselective. We make the hypothesis that pastures can be differentiated from hay meadows based on vegetation indices temporal profiles. A pixel-based classification of S2 image time series is applied to differentiate the two management practices. The classification is then combined with LPIS parcel delineation and high-resolution ancillary data to retrieve homogeneous management units.

Finally, the third specific objective of this thesis is to evaluate the variability in growth status and forage yield between management practices and regions. Grassland outputs (i.e. forage yield and nutritional quality) are strongly related to inputs, i.e. mowing, grazing, and fertilization practices. These relationships are however complex and change depending on pedo-climatic conditions. In this last part, regression models are developed to retrieve grassland biomass yield, nitrogen concentration, canopy nitrogen content, and nitrogen nutrition index from S2 spectral bands and indices. Using these models in combination with the management units and mowing dates retrieved in previous chapters, the spring growth status and the harvested forage yield of different management classes are compared in three agroecological regions.

Outline

This thesis is mainly built from a collection of articles (published or in preparation) addressing the different aspects of GUI monitoring mentioned in section 1.4. Chapter 2 focuses on mowing detection with Sentinel-1 and chapter 3 presents a mowing detection method combining Sentinel-1 and Sentinel-2. In chapter 4, we develop a classification method to retrieve homogeneous management units of pastures and hay meadows. The retrieval of biomass and nitrogen content from Sentinel-2 data is addressed in chapter 5. Finally, general conclusions and perspectives are presented in chapter 6.

Supporting publications

De Vroey, M., Radoux, J. & Defourny, P. (2021). Grassland mowing detection using sentinel-1 time series: potential and limitations. *Remote Sensing*, 13(3), 348. 10.3390/rs13030348

De Vroey, M., De Vendictis, L., Zavagli, M., Bontemps, S., Heymans, D., Radoux, J., Koetz, B., & Defourny, P. (2022). Mowing detection using Sentinel-1 and Sentinel-2 time series for large scale grassland monitoring. *Remote Sensing of Environment*, 280, 113145. 10.1016/j.rse.2022.113145

De Vroey, M., Radoux, J. & Defourny, P. (2023). Classification of grassland management units using optical and microwave remote sensing *Remote Sensing*, 15, 181. 10.3390/rs15010181

De Vroey, M., Radoux, J., Farinelle, A., & Defourny, P. Retrieving grassland nitrogen nutrition status and forage yield from Sentinel-2. In preparation.

Conference papers - talks - posters

De Vroey, M. (2019). Assessing Grassland Use Intensity using Sentinel Optical and SAR data for Biodiversity Assessment. Poster at Living Planet Symposium 2019.

De Vroey, M., Radoux, J., & Defourny, P. (2020). Monitoring land use intensity and ecological value of open biotopes using Sentinel-1 and Sentinel-2. Oral presentation at National Symposium for Applied Biological Science.

De Vroey, M., Radoux, J., Zavagli, M., De Vendictis, L., Heymans, D., Bontemps, S., & Defourny, P. (2021). Performance Assessment of the Sen4CAP Mowing Detection Algorithm on a Large Reference Data Set of Managed Grasslands. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, *IEEE*, 743-746. 10.1109/IGARSS47720.2021.9553269

De Vroey, M., Radoux, J., & Defourny, P. (2022). Grassland types mapping using Sentinel-1 and Sentinel-2. Poster at Living Planet Symposium 2022.

Chapter 2

Grassland Mowing Detection using Sentinel-1 Time Series: Potential and Limitations

This chapter is adapted from the following article:

De Vroey, M., Radoux, J. & Defourny, P. (2021). Grassland mowing detection using sentinel-1 time series: potential and limitations. *Remote Sensing*, 13(3), 348. 10.3390/rs13030348

Abstract

The timing and frequency of mowing events are major aspects of grassland use intensity, which have an impact on their ecological value as habitats. Previous studies highlighted the feasibility of detecting mowing events based on remote sensing time series, most of which used optical imagery. Regular temporal coverage is crucial for accurate mowing detection. This can be an issue when using optical data in cloudy regions. In this chapter, Sentinel-1 C-band microwave data is used for detecting mowing events in various agricultural grasslands. Several mowing detection methods, based on SAR backscattering and interferometric coherence time series, are developed, evaluated, and compared using a large and diverse field data set for training and validation. Results show that 54% of mowing events could be detected in hay meadows, based on coherence jumps. Grazing events were identified as a major confounding factor, as most false detections were made in pastures. Parcels with one mowing event in the summer were identified with the highest accuracy (71%). Overall, this study demonstrates that mowing events can be detected through Sentinel-1 coherence. However, the performances could probably be further enhanced by discriminating pastures beforehand and combining Sentinel-1 and Sentinel-2 data for mowing detection.

2.1 Introduction

Mowing plays a major role in European grassland management. Grasslands are mown, mostly for forage production, but also for ecological conservation (Savage et al., 2021). The timing and frequency of mowing events are major aspects of grassland use intensity and have significant impacts on grassland provisioning and regulating ecosystem services and biodiversity (Humbert et al., 2012; Van Vooren et al., 2018; Tälle et al., 2018; Johansen et al., 2019; Savage et al., 2021). These impacts need to be further studied to adapt grassland management plans to maximize synergies and limit trade-offs between ecosystem services. Precise and exhaustive data on grassland mowing dates and frequency are, however, rarely available to wide extents. Therefore, the automated detection of mowing events through remote sensing would present a great asset.

Previous studies have successfully detected mowing events through sudden decreases in spectral vegetation indices derived from optical imagery and related to biomass, such as the NDVI (Estel et al., 2018; Kolecka et al., 2018; Griffiths et al., 2020; Schwieder et al., 2022; Gómez Giménez et al., 2017). To detect mowing events, it is however crucial to ensure a regular and high observation frequency, which can be a challenge when using optical sensors in cloudy regions (Sano et al., 2007). In that context, Synthetic Aperture Radar (SAR) imagery presents great potential, since active radar sensors are independent of sunlight and cloud cover, making them almost all-weather systems, unlike optical sensors (Ali et al., 2016a; Howison et al., 2018).

A few studies have evaluated the feasibility of detecting mowing events through changes in microwave backscattering (Schuster et al., 2011; Taravat et al., 2019) or interferometric coherence time series (Tamm et al., 2016; Zalite et al., 2016). Although providing encouraging results, studies on grassland mowing detection using SAR time series have mainly been carried out on rather limited study areas and lack sufficient validation data.

In this chapter, the goal is to differentiate grassland mowing dynamics using a mowing detection method based on microwave time series. We assess the potential of Sentinel-1 C-band SAR for detecting mowing events in various agricultural grasslands in a statistically sound manner, using a large and diverse reference data set. The aim

is to answer three questions, namely (1) how accurately can mowing events be detected based on Sentinel-1 time series only, (2) what are the main limitations and confounding factors to mowing detection with Sentinel-1, and (3) can Sentinel-1 based mowing detection be used to identify different grassland mowing dynamics? For this purpose, an intensive field campaign was carried out, providing timely information on mowing and grazing events on various grassland parcels. This allowed several methods to be developed and calibrated, based on the literature and preliminary observations, and to carry out thorough performance analysis, providing a statistically significant evaluation of Sentinel-1's mowing detection potential.

2.2 Study Area

The main study area of this thesis is located in Wallonia, the southern region of Belgium (Figure 2.1). Permanent grasslands cover 35% of the UAA of the Walloon region (Statbel, 2020). The number of parcels and the intensity of management vary between the different agroecological areas of the region (delineated in Figure 2.1). The delimitation of these areas is based on agroecological conditions and cropping systems. The low and relatively flat northern areas of Wallonia (Silty plateaux of Hainaut and Brabant and Hesbaye) are characterized by large cereal and beet crops and few grasslands. Pays de Herve, in the northeast, presents characteristic features of bocage landscapes with very productive pastures and hay meadows. In the Condroz, the topography becomes more rugged and grasslands more frequent. The Fagne-Famenne is dominated by forest and meadows and most of the agricultural area of Ardennes and Lorraine is occupied by pastures and more extensive meadows. The present study encompasses a significant diversity of agro-ecological and farming conditions as it includes Condroz, Fagne-Famenne, and Ardennes, with respectively about 37%, 71%, and 88% of the UAA occupied by grasslands (Statbel, 2015). The other agricultural areas dominated by grasslands (Pays de Herve and Lorraine) were not covered in this study for field campaign feasibility reasons.

Most grasslands in Wallonia are managed through relatively intensive mowing, grazing, or a combination of both (i.e. EUNIS class

"*Permanent mesotrophic pastures and aftermath grazed meadows*"). Extensive hay meadows that are strictly mown without grazing activity (i.e. *"Low and medium altitude hay meadows"*) have become scarcer in recent decades (DEMNA, 2010). In most common grasslands, exploitation activities (grazing or mowing) start in mid-April. In grasslands of high biological interest, supported by the EU CAP, mowing is only allowed after the 16^{th} of June, for flowering purposes, and before the 31^{th} of October.

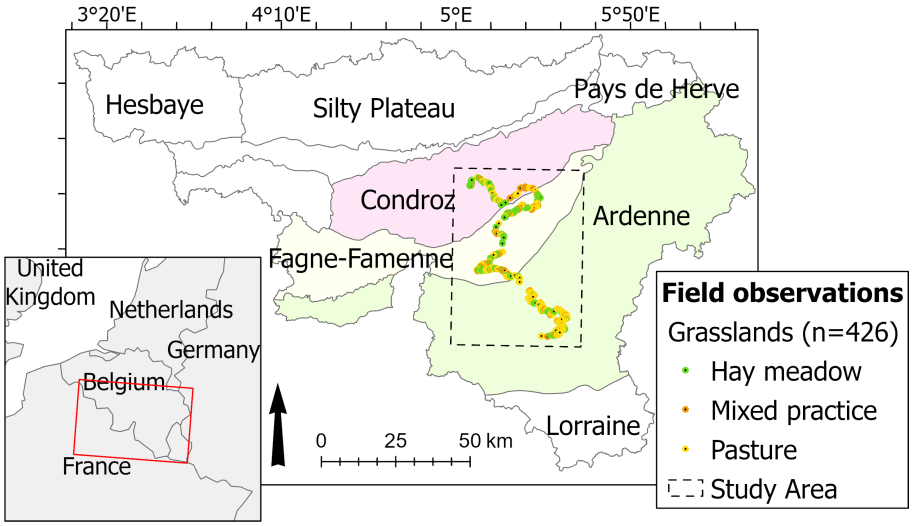

Figure 2.1: Study area. Location of the grassland parcels monitored during the windshield survey across 3 main agroecological regions of Wallonia (Condroz, Fagne-Famenne, and Ardennes). The set of 426 observed parcels includes hay meadows, pastures, and mixed practices.

2.3 Data

2.3.1 Field campaign

An intensive field campaign was carried out to collect an exhaustive reference dataset on grassland management practices. From the 9^{th} of April to the 19^{th} of July 2019, 426 permanent grassland parcels, including pastures, extensive hay meadows, and mixed practices, were

monitored through a windshield survey (Figure 2.1). Each parcel was observed 11 times during the study period. The field visits were carried out with intervals of 6, 12, or 18 days in order to match as much as possible Sentinel-1 A and B combined revisit cycles, with the highest frequency in May and June, when the first cuts are expected to occur and the regrowth would be relatively fast (Figure 2.2). On each visit, the management status of each grassland parcel was recorded ('growing', 'recently cut', 'being cut', 'grazed'). This resulted in a time series of field observations for each parcel. Three types of intervals were defined based on subsequent field observations:

- Mowing intervals, corresponding to the time between the last observation marked as 'growing' (start date) and a 'recently cut'/'being cut' observation (end date) (Figure 2.3 (a));
- Grazing intervals, corresponding to a 'grazed' observation (end date), preceded by any other observation (start date) (Figure 2.3 (b));
- No activity intervals, corresponding to any observation (start date) followed by an observation marked as 'growing' (end date) (Figure 2.3 (c)).

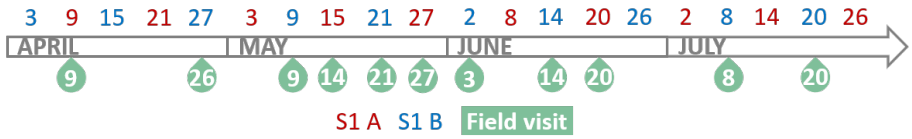

Figure 2.2: Sentinel-1 A and B (descending) acquisition dates over the study area and corresponding field observation dates between April and July 2019.

This survey resulted in a total of 4260 observation intervals (10 intervals for each of the 426 parcels) including 261 mowing intervals.

Based on observations, four types of grasslands could be differentiated: parcels where only grazing was observed (pastures, n=201), grazed parcels with at least one mowing event (mixed practices, n = 61), and parcels without grazing, but with at least one mowing event

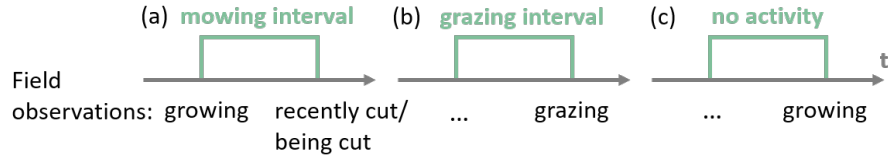

Figure 2.3: Schematic representation of the reference intervals derived from the field observations.

(hay meadows, n = 154). On 10 parcels, the management practice could not be defined with certainty based on field observations. These parcels were discarded from the reference dataset.

Approximately half of the observed parcels were randomly selected for calibration (n = 220) and the remaining parcels were kept for validation (n = 196). The selection was stratified to ensure that the three types of grasslands were represented in each subset.

2.3.2 Sentinel-1 Time Series

Sentinel-1 data was acquired in Single Look Complex (SLC) format and interferometric wide (IW) swath mode with dual polarization (vertical transmission with vertical reception (VV) and horizontal reception (VH)). The processing to Ground Range Detected (GRD) backscattering coefficient (γ^0) and interferometric coherence with a resolution of 15 m was performed using the Sentinel-1 Toolbox of the Sentinel Application Platform (SNAP version 6.0). The processing chains shown in Figure 2.4 include calibration, georeferencing, deburst, InSAR coherence estimation (with an averaging window size of azimuth × range: 3 × 10), and terrain correction. Both chains were run on all acquisitions from Sentinel-1 descending pass covering Wallonia during the study period in 2019. Only descending pass acquisitions were treated in this study because coherence should not be computed between different passes (ascending or descending), since the look direction varies. Moreover, field visits were carried out with a minimum interval of 6 days, matching just descending pass acquisition dates, for practical and timing reasons. Coherence was computed between consecutive Sentinel-1 A and B images, to reach a 6-day revisit cycle instead of 12, as suggested by Tamm et al. (2016).

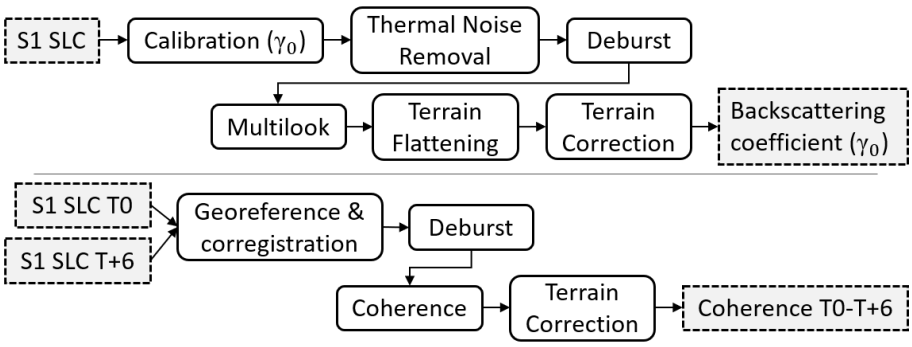

Figure 2.4: Schematic overview of the Sentinel-1 processing chain from single look complex (SLC) interferometric wide (IW) swath mode with dual polarization (VV, VH) to Ground Range Detected (GRD) backscattering coefficient (γ^0) and interferometric coherence with a 15 m resolution. The processing was performed using the Sentinel-1 Toolbox of ESA's SNAP (version 6.0).

SAR imagery is characterized by an inherent variance (speckle) caused by constructive and destructive interference between randomly distributed scatterers within a pixel (Lee et al., 1994). This speckle effect can be observed on apparent homogeneous surfaces, such as herbaceous covers. To spatially smooth the SAR data, the regional LPIS layer (Milionis, 2016) — established on an annual basis by the Walloon Administration for the EU Common Agriculture Policy and referencing the location, the extent, and the main crop of agricultural parcels in Wallonia — was used to average the backscattering and coherence signal per parcel. A 10 m inner buffer was applied to the parcel polygons before averaging, to reduce border effects.

2.4 Methods

Several mowing detection methods were tested based on previous studies and time series observation and compared to each other. The methods are based on change detection in time series of a backscattering coefficient on the one hand and of interferometric coherence on the other. All methods were tested using VV and VH signal, as well as a combination of both polarizations, namely the ratio (VV/VH) in the

case of backscattering and the average (mean(VV,VH)) in the case of coherence.

2.4.1 Backscattering Detection Method

This first method, based on γ^0 backscattering time series, consists of detecting the occurrence of a signal increase followed by a signal decrease, with respective threshold magnitudes, which have been observed after mowing events in several studies (Bargiel et al., 2010; Schuster et al., 2011; Curnel, 2015). The change in backscattering after a mowing event could be related to a stronger contribution of soil surface scattering, compared to volume scattering of vegetation in tall grass. The method selected here is largely based on the two axioms used by Schuster et al. (2011), namely (1) the occurrence of a signal increase followed by a decrease and (2) the magnitude of those changes above a fixed threshold. For each parcel extracted time series, each value γ_i is compared to the previous γ_{i-1} and the next γ_{i+1}. Two conditions must be met to detect a mowing event.

(1) The value γ_i needs to be larger than the previous value γ_{i-1} and the next value γ_{i+1}:

$$\gamma_{i-1} < \gamma_i > \gamma_{i+1} \tag{2.1}$$

(2) The amplitude of the changes, expressed as percent increase D_{up} and decrease D_{down}:

$$D_{up} = 100 * (\frac{\gamma_i - \gamma_{i-1}}{\gamma_i}) \tag{2.2}$$

$$D_{down} = 100 * (\frac{|\gamma_{i+1} - \gamma_i|}{\gamma_i}) \tag{2.3}$$

must be higher than the average percent increase and decrease by a fixed threshold (K_{up}% and K_{down}%, respectively). The average percent change describes the signal variation along the entire time series and is computed for each parcel, to be used in the following equations:

$$D_{up} - \frac{\sum_{i=0}^{N} \frac{|\gamma_i - \gamma_{i-1}|}{\gamma_i} * 100}{N} \geq K_{up}\% \tag{2.4}$$

and

$$D_{down} - \frac{\sum_{i=0}^{N} \frac{|\gamma_{i+1} - \gamma_i|}{\gamma_i} * 100}{N} \geq K_{down}\% \tag{2.5}$$

A detection is made when Equations (2.1), (2.4) and (2.5) are true. This method was tested with different threshold values K_{up}% and K_{down}% (Table 2.1).

Table 2.1: Overview of the considered mowing detection methods. The backscattering methods were tested with different input polarizations, percent increase thresholds (K_{up}%), and different ratios between percent decrease and increase thresholds (K_{down}/K_{up}). The coherence jump detection methods were tested with different polarizations, smoothing approaches, window sizes (d), and absolute thresholds (a.t.,k) or relative threshold (r.t.) parameters (α).

Input	K_{up}%	K_{down}/K_{up}	
γ_0 (VV, VH, ratio)	2; 5; 10; 15; 20; 25	1/4 ; 1/2 ; 3/4 ; 1	
Input	**Smoothing**	d	**Detection Threshold**
Coherence (cohVV; cohVH; cohVVVH)	Mean shift	7 ; 9 ; 11	Absolute (a.t) k= {0.5, 0.25, 0.1, 0.075, 0.05, 0.025, 0.01, 0.005, 0.0025} ;
	Linear regression	3 ; 4 ; 5 ; 6	Relative (r.t.) α= {0.0005, 0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.2, 0.4}
	Two means	8 ; 10	p-value = {0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.25, 0.5, 0.75}

2.4.2 Coherence Jump Detection Methods

The second set of methods is based on the observation of a high jump in coherence time series right after a mowing event. A mowing event implies a drastic reduction of above-ground biomass. According to observations made in previous studies (Zalite et al., 2016; Tamm et al., 2016), such a reduction in biomass causes a sudden increase from relatively low coherence values to higher ones for a given period after the mowing event, before the grass grows tall again. Since coherence tends

to fluctuate between consecutive dates despite spatial filtering, temporal smoothing was applied to reveal significant increases. Based on preliminary observations of coherence time series on various parcels, three jump detection methods were considered that involve different smoothing approaches (Figure 2.5).

Mean shift: The first method, called mean shift, consists of a sliding averaging window of a maximum size d with an adjusting symmetry. It allows to smooth the signal while highlighting significant shifts in the temporal mean such as jumps. When applying the mean shift sliding window, each raw value T_i is replaced by the average of $[T_{i-r_1}, T_{i+r_2}]$, resulting in a smoothed value $ms(T_i)$. The number of points before (r_1) and after (r_2) included in the window are adapted to the local variance of the signal in each window. They are defined by the following limits (Equation (6)) that result in a maximum window size $d = 2 * r_{max} + 1$.

$$\begin{cases} r_1 = r_{max} \\ r_2 = [0, r_{max}] \end{cases} \quad OR \begin{cases} r_1 = [0, r_{max}] \\ r_2 = r_{max} \end{cases} \tag{2.6}$$

Between those limits, the window symmetry parameters r_1 and r_2 adjust for each window to minimize the standard error of the average (σ_x):

$$\sigma_x = \frac{\sigma_{[T_{i-r_1}, T_{i+r_2}]}}{\sqrt{r_1 + r_2 + 1}} \tag{2.7}$$

where σ is the standard deviation of $[T_{i-r_1}, T_{i+r_2}]$. Jumps are detected between smoothed values when the difference $ms(T_i) - ms(T_{i-1})$ exceeds a fixed threshold. The method was tested with absolute thresholds (a.t.) (k) or relative thresholds (r.t.) to the standard error. In the second case, a Student's test (t-test) with a given significance level α is performed. All the values tested for the maximum window size d and the thresholds k and α are summarized in Table 2.1.

Linear regression: In the second method, the time series are smoothed by linear regression using an asymmetric sliding window. Due to the growth of vegetation before the mowing event, the coherence is expected to gradually decrease and then suddenly increase after the mowing event. In this method, each raw value T_i is compared to the previous smoothed value $lr(T_{i-1})$ obtained by linear regression of

$[T_{i-d}, ..., T_{i-1}, T_i]$. This allows to consider a potential slope in the coherence profile and detect sudden increases, compared to the previous signal trend. A detection is made when the difference $T_i - lr(T_{i-1})$ exceeds a given threshold. This method was also tested with different window sizes d, and absolute (k) and relative (α) thresholds (Table 2.1).

Two means: The third jump detection method is based on an even-size sliding window d in which a statistical hypothesis test verifies if there is a significant change in temporal mean coherence in the middle of the window. This is another way to detect a sudden increase in a coherence time series. The null hypothesis H_0 is the approximation by the average value of the entire window $mean([T_1 : T_d])$. The alternative hypothesis H_1 is the approximation by a "step model" consisting of two different averages for the first half $[T_1 : T_{d/2}]$ and second half $[T_{d/2+1} : T_d]$ of the window. The F_{stat} used for the test is computed using Equation (2.8):

$$F_{stat} = \frac{\left(\frac{SSR_0 - SSR_1}{k_1 - k_0}\right)}{\left(\frac{SSR_1}{d - k_1}\right)} \tag{2.8}$$

where $SSR_{[0,1]}$ and $k_{[0,1]}$ are, respectively, the Sum of Squared Residuals and the number of parameters of $H_{[0,1]}$. The probability density function of F_{stat} is derived from n, k_0 and k_1 and a p-value is computed for each date. A p-value under a given threshold implies a rejection of H_0. In that case, and if the second mean value is higher than the first, a coherence jump is detected. Table 2.1 gives an overview of all the coherence jump detection methods applied using the different inputs, window sizes, and detection thresholds.

2.4.3 Method Calibration, Evaluation and Validation

Each mowing detection method was calibrated to determine the best parameter values, i.e. window sizes and detection thresholds. The robustness of these methods was then tested with respect to potential confounding factors related to the use of Sentinel-1 data. The calibrated methods were then validated on an independent set of grassland parcels to estimate the potential of Sentinel-1 based mowing detection for identifying different grassland mowing dynamics.

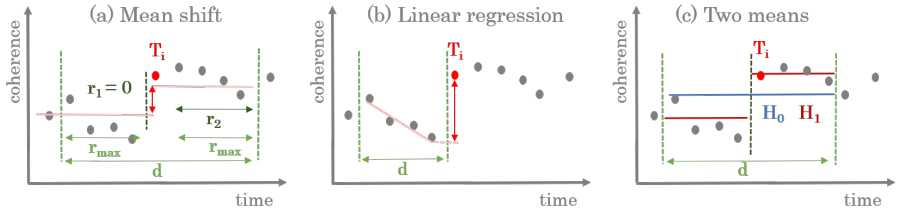

Figure 2.5: Schematic representation of the 3 coherence jump detection methods. A jump occurs at T_i. (**a**) The mean shift method with maximum window size $d = 2 * r_{max} + 1 = 9$ ($r_1 = 0$ and $r_2 = r_{max} = 4$), where the difference (red arrow) between the smoothed values is evaluated, (**b**) the linear regression method with window size $d = 4$, where raw value T_i is compared to the previous smoothed value and (**c**) the two means method of window size $d = 8$, comparing the hypothesis H_0 (1 mean) and H_1 (2 means).

Calibration and robustness analysis

The calibration was performed on a subset of the training dataset with parcels presenting ideal size, shape, and slope orientation. Those ideal parcels have an area larger than 1 ha, a width larger than 30 m and the hill shade has to be above the median of the data set. Hill shade was used as an indicator for slope orientation, as it is based on the orientation of the slope relative to an illumination source angle and shadows. These three requirements correspond to parcel characteristics potentially affecting the Sentinel-1 signal. Small parcels are expected to be more challenging to monitor because of the limited amount of pixels included in the spatial smoothing. The shape of a parcel could be an issue because narrow parcels might be impaired by border effects (mixed pixels). Finally, the orientation of the slope relative to the direction of the radar beam will have an impact on the interaction of the signal with the surface.

The calibrated methods were then applied to parcels presenting less ideal characteristics to observe the impact of size, shape, and slope orientation on the mowing detection accuracy. In addition to parcel characteristics, the robustness of the methods to grazing activities was assessed.

During the calibration and robustness analysis, the accuracy of mowing detection was evaluated by crossing the reference mowing in-

tervals with detection intervals. When a detection interval intersects a reference mowing interval, it is considered as a true positive (TP). If no reference mowing interval overlaps a detection, it is a false positive (FP) and if no detection overlaps a reference mowing interval, it is counted as a false negative (FN). The remaining intervals are true negatives (TN).

Since mowing events are relatively rare, there is a strong imbalance between positives and negatives. Therefore, in addition to the broadly used overall accuracy (OA), the Matthews Correlation Coefficient (MCC) was used for calibration and evaluation, along with the *detection rate* and *precision* of mowing detection. The MCC (Equation 2.9) was first introduced by Matthews (Matthews, 1975) and the metric is particularly suited to measure the quality of an unbalanced binary classification (Boughorbel et al., 2017). MCC values range from −1 to 1, 0 being the equivalent of a random classification. The *detection rate* and *precision* are computed with Equations (2.10) and (2.11).

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{2.9}$$

$$detection\,rate = \frac{TP}{(TP + FN)} \tag{2.10}$$

$$precision = \frac{TP}{(TP + FP)} \tag{2.11}$$

Validation

The best-performing and most robust method was then applied to the independent validation data set. In this case, the performances were evaluated per parcel, instead of per time interval. The aim is to assess the potential of Sentinel-1 based mowing detection methods for identifying different mowing dynamics of grasslands. The assessment was performed for the four grassland management classes observed during the field campaign: (i) unmown pastures, (ii) parcels with 1 spring mowing (before the 20^{th} of June), (iii) with 1 summer mowing (after the 20^{th} of June) and (iv) with 2 mowing events during the study period.

2.5 Results

2.5.1 Method Calibration

All methods were run on the ideal calibration parcels with a range of different detection parameters (Table 2.1). In total, 12 configurations were considered for methods based on backscattering. For coherence jump detection, 48 different configurations of polarization, smoothing approach, window size, and threshold type were considered. The first observation is that the detection method based on backscattering performed very poorly compared to the coherence jump detection methods. The maximum MCC obtained with backscattering is 0.25, using γ^0 VV time series, while the highest MCC values obtained with the different coherence jump detection methods range from 0.26 to 0.49. Therefore, the focus was set on coherence jump detection methods for further analysis.

The different smoothing methods, window sizes and threshold types are compared based on OA, *detection rate*, *precision* and MCC along detection threshold values (Figure 2.6). The OA (unbroken blue line) reaches high values ranging from 92% to 96% for all methods. It drops significantly with low absolute detection thresholds (a.t.), for example, to 45% for k = 0.0025 with the cohVV linear regression method (d = 6), because of the numerous false detections made with such a low detection threshold (Figure 2.6d).

As expected, when the detection threshold increases ($1-\alpha$, $1-p$-value and k), the *precision* increases as fewer false detections occur, but the *detection rate* declines as more mowing events are omitted. No significant differences were observed in the shapes of the *detection rate* and *precision* curves, between the different input polarizations (cohVV, cohVH, and cohVVVH) or window sizes. On the other hand, the shape of the curves varies from one smoothing method to another, as shown in Figure 2.6. Both with absolute and relative thresholds, the threshold value at which *precision* and *detection rate* reach a balance and the curves cross, is higher for the linear regression method than for the mean shift. For the mean shift method (Figure 2.6 (a,b)), when α tends toward 0 or k increases, the *detection rate* drops relatively low as the detections become more precise. Meanwhile for the linear regression method (Figure 2.6 (c,d)), the *detection rate* remains high with more

precise detection thresholds. For the two means smoothing method, the *detection rate* also drops relatively fast while the *precision* does not increase significantly.

The MCC metric (dashed blue line), peaks above 0.4 for all methods except for the two means method (Figure 2.6 (e)). The two-means method was therefore also discarded for further analysis. The MCC peak was used here to select the best-performing methods and detection thresholds. However, even at these optimal thresholds, either the *precision* or the *detection rate* are lower than 50% for all methods.

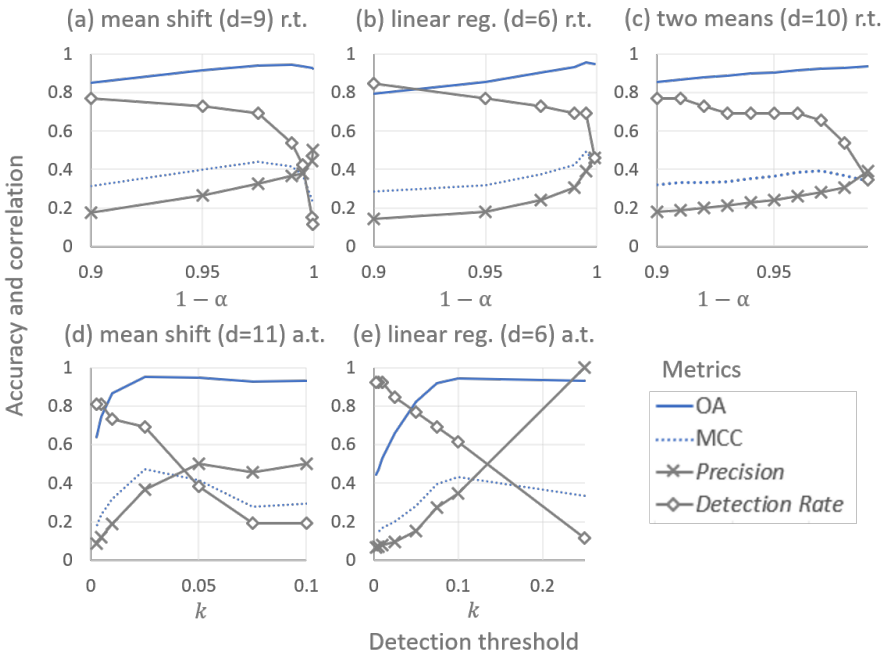

Figure 2.6: Overall Accuracy (OA), Matthews Correlation Coefficient (MCC), *precision* and *detection rate* of several mowing detection methods based on VV coherence, with varying detection threshold parameters. The graphs show the performances of the mean shift and linear regression methods with relative threshold (r.t.) and absolute threshold (a.t.) and of the two means method, with the best window size (d) for each.

For each input feature set (cohVV, cohVH,cohVVVH), the two methods delivering the highest MCC value were selected for further

analysis. The six selected methods are (i) cohVH mean shift ($d = 9$) with relative threshold ($\alpha = 0.001$), (ii) cohVH linear regression ($d = 5$) with r.t. ($\alpha = 0.0005$), (iii) cohVV mean shift ($d = 11$) with a.t. ($k = 0.025$), (iv) cohVV linear regression ($d = 6$) with r.t. ($\alpha = 0.005$), (v) cohVVVH linear regression ($d = 5$) with a.t. ($k = 0.1$) and (vi) cohVVVH linear regression ($d = 6$) with r.t. ($\alpha = 0.005$). The MCC values obtained for these methods range from 0.44 for (ii) to 0.49 for (iv). These calibrated methods are listed in Table 2.2.

2.5.2 Robustness to Confounding Factors

To analyze their robustness to potential confounding factors, the mowing detection methods were tested on less ideal parcels. The *precision* (*pre.*) and *detection rates* (*d.r.*) obtained for ideal, small and narrow parcels, for parcels with lower hill shade values, for all non-ideal parcels, and for all parcels are compared in Table 2.2.

On ideal parcels, the calibration based on MCC metrics resulted in a balance between omissions and false detections, depending on the methods. The cohVH mean shift method shows the least balance, with the highest *precision* (78%) and the lowest *detection rate* (27%). The most balanced calibrated method is cohVVVH linear regression with absolute threshold, with a *precision* and *detection rate* of 42% and 54%, respectively. Both methods with VH coherence were calibrated with a higher *precision*, while the methods with cohVV and cohVVVH resulted in higher sensitivities at the optimal threshold. The highest *detection rate* (69%) on ideal parcels was obtained with both cohVV methods and the cohVH linear regression method with a relative threshold.

Among the parcel characteristics, the most significant impact is observed for narrow parcels, as both *precision* and *detection rate* are significantly lower for all methods. The largest impact of parcel shape is observed with the cohVH mean shift method, which was also the least balanced. The *precision* and *detection rate* of this method respectively dropped to 13% and 8% on the narrow parcels. The individual effects of size and slope orientation on the accuracy of mowing detection are less significant or consistent through the different methods. Some positive impacts are even observed on the *precision*.

To better observe the combined impact of the three parcel characteristics on the *detection rate* and *precision* of mowing detection, ideal and

non ideal parcels are compared (Figure 2.7 (a)). For all methods, the *detection rate* is lower on non-ideal parcels. The effect on the *precision* is, however, more variable. The most robust methods are cohVH linear regression and both cohVVVH linear regression methods, for which the *detection rate* and *precision* vary by less than 10% between ideal and non ideal parcels. For cohVV linear regression, the *precision* is not impacted by the parcel characteristics, but the *detection rate* drops by 20%. As before, the largest impacts are observed on cohVH mean shift, where the *precision* drops by 31%.

In addition to parcel characteristics, the impact of grazing activities on the detection accuracy was evaluated (Figure 2.7 (b)). As expected the *detection rate* does not vary when unmown pastures are removed from the data set. The *precision* is however significantly higher for all methods, namely 64% for cohVVVH linear regression with a.t., 72% for cohVV linear regression, and 86% for cohVH linear regression, while these methods have *precisions* of 43%, 42% and 57%, respectively, when considering all types of grasslands. This means that a large part of the false detections was due to unmown pastures. When the mixed practices are removed and only the hay meadows are considered, the *precision* and *detection rate* barely change and no consistent impact is observed through the different methods.

The most important increase of *precision* (+30%) after discarding unmown pastures was observed for the cohVV linear regression method (d = 6) with r.t. (α = 0.005), which has a reasonable *detection rate* of 54% on all parcels and reached a 69% *detection rate* on parcels with ideal characteristics. It was therefore selected as the overall best-performing method and used for further parcel-based validation.

2.5.3 Parcel-Based Validation

The calibrated cohVV linear regression (d = 6) method with r.t. (α = 0.005) was applied on the independent validation data set, for grassland management practice classification (Figure 2.8). The aim is to assess the performances for identifying different mowing dynamics of grasslands, based on Sentinel-1 mowing detection.

Of the 102 pastures, 54% were correctly identified as unmown (first bar of Figure 2.8), which means that false detections were made in

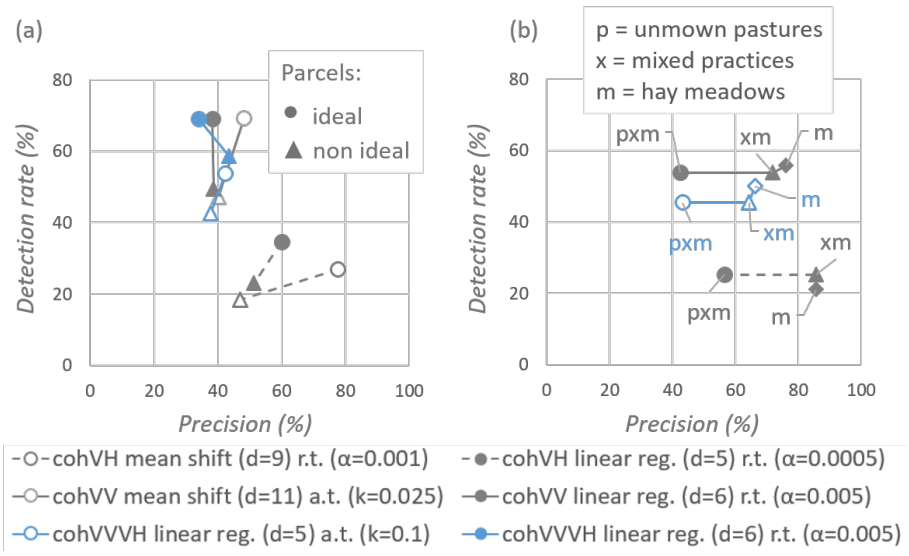

Figure 2.7: Robustness of Sentinel-1 coherence mowing detection methods to parcel characteristics (size, slope orientation, and shape) and grazing activities. Graph (**a**) shows the changes in *precision* and *detection rate* between detections on large, broad and well oriented parcels (ideal) and on parcels with less ideal characteristics (non ideal). Graph (**b**) shows the changes in *precision* and *detection rate* when removing unmown pastures (p) and then mixed practices (x), considering hay meadows (m) alone.

almost half of the pastures. This confirms the results of the robustness analysis, which identified grazing as a major confounding factor to mowing detection with Sentinel-1.

The second bar of Figure 2.8 shows the performances on the parcels with one spring mowing. Here too, 54% were correctly identified. In the remaining spring mowing parcels, the mowing event was omitted and in 14% a false detection was made in addition to the omission. Much better results were obtained with summer mowings (third bar of Figure 2.8), as 71% of the parcels with one summer mowing were correctly identified. A second false detection was made in 15%, and the summer mowing event was omitted in 14%, in two of which a false detection was made in the spring.

Finally, from the parcels with two mowing events (fourth bar of Figure 2.8), only 32% were entirely correct. Here as well, most of the summer events were detected (64%), while many spring mowings were omitted (48%). Both mowing events were omitted in only four parcels, one of which also counted as a false detection.

Considering the whole validation data set, 56% of all parcels and 59% of the mown parcels were correctly identified in terms of mowing dynamics, based on Sentinel-1 coherence mowing detection. Most errors consist of false detections in grazed parcels and omitted spring mowing events.

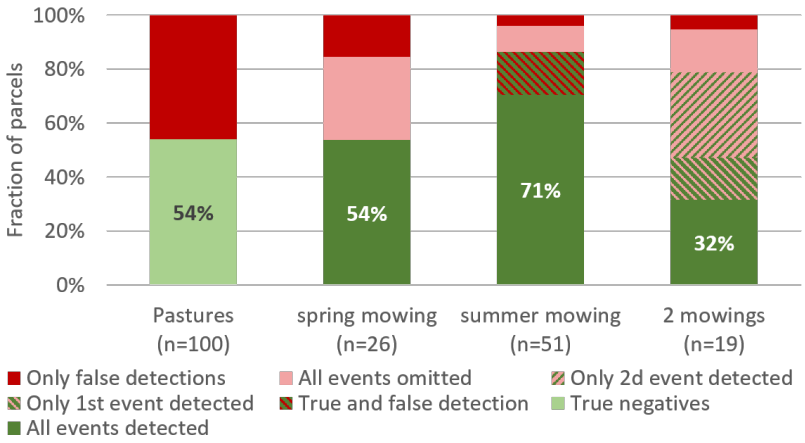

Figure 2.8: Per parcel evaluation of a Sentinel-1 VV coherence mowing detection method to assess its potential to identify different mowing dynamics. For each type of dynamic, the fraction of correctly estimated parcels is given, along with the types of errors that occur (false detections, omissions, or both).

Table 2.2: Impact of the size, topography, and shape of the parcels on the accuracy of calibrated mowing detection methods, based on VH coherence (cohVH), VV coherence (cohVV), and the average of VV and VH coherence (cohVVVH) with absolute (a.t.) and relative (r.t.) threshold and window sizes d. The threshold parameters α and k were optimized based on the Matthews Correlation Coefficient (MCC), to obtain a balance between *precision* (*pre.*, %) and *detection rate* (*d.r.*, %).

	Ideal	**Small**	**Slope Dir**	**Narrow**	**Non Ideal**	**All**
n	590	430	770	330	1470	2200
cohVH mean shift (d = 9) r.t. (α = 0.001)						
pre.	78	50	53	13	47	57
d.r.	27	8	20	8	18	22
cohVH linear reg. (d = 5) r.t. (α = 0.0005)						
pre.	60	63	59	33	51	54
d.r.	35	21	31	15	23	25
cohVV mean shift (d = 11) a.t. (k = 0.025)						
pre.	48	35	44	25	40	42
d.r.	69	63	59	31	47	53
cohVV linear reg. (d = 6) r.t. (α = 0.005)						
pre.	38	32	44	32	39	39
d.r.	69	42	57	46	49	54
cohVVVH linear reg. (d = 5) a.t. (k = 0.1)						
pre.	42	32	44	26	38	38
d.r.	54	63	45	46	43	45
cohVVVH linear reg. (d = 6) r.t. (α = 0.005)						
pre.	34	46	49	30	44	41
d.r.	69	50	67	46	59	61

2.6 Discussion

The results of this study show the potential of grassland mowing detection using InSAR coherence, as already suggested by a study on a limited number of grassland parcels (Tamm et al., 2016). Overall, it was shown that mowing events can be detected by identifying jumps in coherence time series. Based on a large and diverse reference data set, several methods could be developed, calibrated, and thoroughly analyzed on different types of grassland. Once calibrated, most methods showed similar performances and robustness. Nevertheless, the mowing detections based on VV coherence smoothed by linear regression with a window d = 6 and a relative threshold ($\alpha = 0.005$) were slightly more precise and accurate, especially on hay meadows.

2.6.1 Mowing Detection Methods

The methods based on the hypothesis of the occurrence of an increase and subsequent decrease in backscattering after a mowing event performed poorly during the calibration phase. Schuster et al. (2011) accurately detected all mowing events in two semi-natural meadows, using TerraSAR-X σ^0 images, while this study used Sentinel-1 C-band γ^0. Their mowing detection method should therefore be further tested with X-band SAR data on a larger and richer data set.

In this study, it was however shown that mowing events can be detected by identifying jumps in smoothed coherence time series. The 'two means' smoothing method could be discarded based on lower MCC values, but performances varied little between the different input polarizations, window sizes, and detection threshold types for the mean shift and linear regression methods (Table 2.1). In ideal cases, the mowing event causes such an ample coherence increase that it is easily detected, whichever method. On the contrary, regardless of the method and parameters, some mowing events cannot be detected at all, and coherence jumps can be caused by signal noise or other surface changes. As an illustration, eight coherence time series extracted on different types of parcels, along with the mowing events observed on the field are shown in Figure 2.9. Therefore, most errors are independent of the detection method but inherent to the signal and its interactions with other factors. To understand the causes of these limitations,

a robustness analysis was carried out to identify the confounding factors to mowing detection with Sentinel-1 coherence.

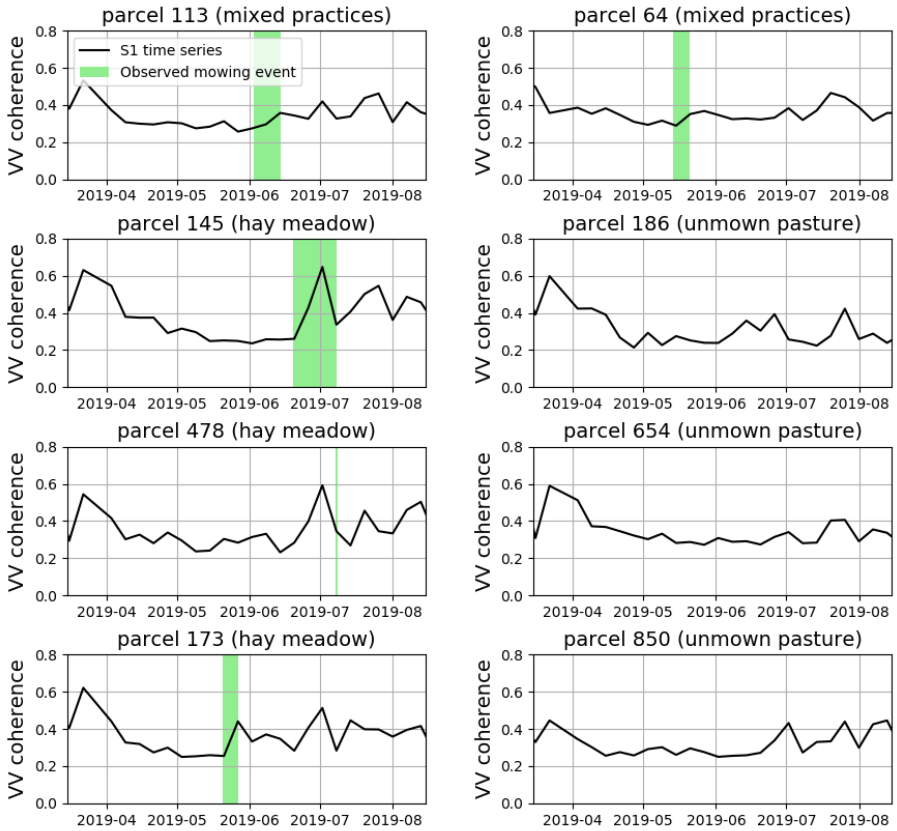

Figure 2.9: Sentinel-1 VV coherence time series extracted from a selection of 8 grassland parcels (2 mixed practices, 3 hay meadows, and 3 unmown pastures). The green areas represent time intervals in which a mowing event occurred according to field observations. The green line represents a mowing event observed on the day of a field visit.

2.6.2 Limitations

The *detection rate* and *precision* metrics in Table 2.2 and Figure 2.7 show that the size, shape and slope orientation of the parcels can hinder the

detection of mowing events for most methods. Fewer mowing events are detected in small or narrow parcels or parcels with a less ideal slope orientation and with some methods, more false detections are made in these less ideal parcels. The differences in accuracy are, however, minor and not always consistent throughout the methods.

Of the confounding factors tested in this study, grazing is the most important one. Many false detections were made in parcels that were not mown but grazed during the study period. During validation, false detections were made in almost half of the unmown pastures, confirming that grazing is a major confounding factor for mowing detection. As coherence increases with decreasing grass height and biomass (Zalite et al., 2016), a grazing event with a large stock density could result in a coherence jump, similar to a mowing event.

In hay meadows and mixed practice parcels, the coherence jump detection method developed in this study allowed to detect most summer mowings, while about half of the spring mowings were omitted (Figure 2.8). This could be explained by a slower regrowth of the grass after summer cuts, causing the coherence to remain high for a longer period and making the detection easier.

The confounding factors assessed in this study explain some of the false detections and missed mowing events. However, in some cases, the cause of errors remains unclear. Other factors could hinder the detection of mowing events by impacting the radar signal, such as the presence of trees and shrubs or the water content of the soil and plant. The impact of water content on radar signal has already been shown previously (Tampuu et al., 2020). It was also shown to be a potential confounding factor to mowing detection (Curnel, 2015; Tamm et al., 2016). In terms of agricultural practices, swaths of grass left on the parcel to dry could also be a confounding factor that alters the signal response after a mowing event.

The mowing detection method developed in this study is object-based and relies on the availability of grassland parcel boundaries for spatial smoothing. The regional LPIS provided this information for this study area. The LPIS parcels are based on farmer declarations and in reality, they are often divided into smaller management units. Mowing events on smaller management units would be omitted when averaging the signal over the larger declared parcels. Alternatively, parcel bound-

aries could be retrieved by automatic delineation based on a convolutional neural network (CNN) (Waldner and Diakogiannis, 2020).

2.6.3 Reference Data and Quality Metrics

The major limiting factor in the development and evaluation of mowing detection methods is the availability of precise and complete field data (Reinermann et al., 2020). The large field campaign carried out in the framework of this study provided information on the land use status of more than 400 agricultural grassland parcels across different agroecological regions of Wallonia, on Sentinel-1 acquisition dates. Thanks to this rich data set, it was possible to statistically evaluate the mowing detection potential of Sentinel-1 in agricultural grasslands.

The choice of adequate quality metrics that properly show the performances of a detection method with respect to the context and overarching goal, is another critical point. In the case of the detection of a rare event, the classes of occurrence and absence of the event are unbalanced. Due to this disproportion, the widely used overall accuracy (OA) is mainly influenced by the numerous absences and is not representative of the detection accuracy. Indeed, the OA reached values above 90% for all methods during calibration, while the *detection rate* and *precision* were relatively low (Figure 2.6). The *detection rate* and the *precision* focus on the true positives, false positives, and false negatives and represent the capacity of a method to detect rare occurrences. Increasing the detection threshold enhances the *precision*, but reduces the *detection rate*, as more events are omitted. When calibrating the mowing detection methods, a compromise needs to be made between maximizing the *detection rate* and minimizing the false detections. In this study, the MCC metric was used to calibrate the methods and select the optimal detection thresholds, as it is well adapted for unbalanced binary classifications (Boughorbel et al., 2017). Depending on the methods, this resulted in a higher *precision* or a higher *detection rate* and more or less balance between both metrics. In practice, the detection threshold needs to be set according to the context and the motivation for mowing events detection.

2.6.4 Potential and Prospects

The Sentinel-1 coherence mowing detection method developed in this study allows to detect most summer mowings in hay meadows. The combined use of Sentinel-1 A and B, ensuring a 6-day revisit cycle and therefore a 6-day baseline for the coherence, was a great asset. With a single satellite, providing images on a 12-day basis, more events would be omitted since the coherence would not increase as much after a mowing event, due to the regrowth of vegetation between subsequent acquisitions. This would significantly lower the *detection rate*, especially in the spring. On the other hand, it is likely that 3 satellites, combined for a 4-day revisit time, would lead to better results.

However, even with both satellites, considering the diversity of grassland types and management dynamics, the mowing detection lacks precision and some mowing events cannot be detected based only on Sentinel-1 coherence. To improve precision, it would be a great asset to differentiate grazed parcels from hay meadows beforehand, as most false detections were made in pastures. Previous studies have discriminated pastures from hay meadows, based, for example, on backscattering coefficients (Curnel, 2015) or vegetation indices derived from optical data (Gómez Giménez et al., 2017).

In the prospects of grassland use intensity assessment and habitat monitoring, it is critical to be able to detect early spring mowing events with more certainty, to differentiate more intensively managed grasslands from extensive grasslands with a higher ecological value. Performances could probably be improved by combining the Sentinel-1 coherence mowing detection, calibrated to obtain high precision, with an optical method based on the detection of decreases in NDVI (Estel et al., 2018; Kolecka et al., 2018).

2.7 Conclusions

In this chapter, the full potential of Sentinel-1 for detecting mowing events was assessed, by estimating the accuracy of mowing detection, identifying the main limitations and confounding factors, and finally evaluating the potential for differentiating grassland mowing dynamics. Using timely field data on the state and the use of 426 grassland

parcels, four Sentinel-1-based mowing detection methods were evaluated with various parameter configurations on diverse grasslands. The methods were based on two main observations of SAR signal response to a mowing event, namely the occurrence of a backscattering signal rise and decrease and an interferometric coherence jump. While the methods based on backscattering showed very low performances, the coherence jump detection methods allowed to accurately detect a majority of mowing events in hay meadows. However, due to the omission of a significant amount of spring mowings and many false detections in pastures, only 56% of grasslands in a fully independent data set was correctly identified in terms of mowing dynamics. The size, shape topography, and agricultural practices explain some of the errors, but a part remains uncertain and must be due to other signal interactions. The results of this study confirm that it is feasible to detect mowing events based on coherence jumps. Performances could probably be further enhanced by discriminating pastures beforehand and combining Sentinel-1 and Sentinel-2 data. However, a compromise will always have to be made between *detection rate* and *precision* and the detection threshold needs to be set accordingly, depending on the context and the purpose of mowing detection.

Chapter 3

Mowing detection using Sentinel-1 and Sentinel-2 time series for large scale grassland monitoring

This chapter is adapted from the following article:

De Vroey, M., De Vendictis, L., Zavagli, M., Bontemps, S., Heymans, D., Radoux, J., Koetz, B., & Defourny, P. (2022). Mowing detection using Sentinel-1 and Sentinel-2 time series for large scale grassland monitoring. *Remote Sensing of Environment*, 280, 113145. 10.1016/j.rse.2022.113145

Abstract

Precise information on mowing dynamics is essential to understand the impact of Management practices on grassland ecosystem services but is rarely available at large scale and with sufficient temporal and spatial resolution. In the previous chapter, the potential and limitations of mowing detection with C-band microwave data were assessed. In this chapter, we build on previous findings and present a new method combining the regular observations of S1 and the better accuracy of S2 grassland mowing detection algorithms. This multi-source approach for grassland monitoring was assessed over large areas and in various contexts. The method was first validated in six European countries, based on Planet image interpretation. Its performances and sensitivity were then thoroughly assessed in an independent study area using a more precise and complete reference dataset based on the intensive field campaign carried out in Wallonia. Results showed the robustness of the method across all study areas and different types of grasslands. The method reached an F1-score of 79% for detecting mowing events on hay meadows. Furthermore, the detection of mowing events along the growing season allows to classify mowing dynamics with an overall accuracy of 69%. This is promising for differentiating grasslands in terms of management practices. The method could therefore be used for large-scale grassland monitoring to support agri-environmental schemes in Europe.

3.1 Introduction

Precise information on mowing dynamics is essential to understand the impact of management practices on grassland ecosystem services. However, such data are rarely available on a broad scale and with sufficient temporal and spatial resolution. Therefore, automated mowing detection based on remote sensing can be a great asset.

In the previous chapter, the potential and limitations of automated grassland mowing detection based on Sentinel-1 data were evaluated. The results showed that mowing events could be detected, based on jumps in interferometric coherence time series. Many events were however omitted, especially in the spring. Moreover, the precision was very low, mainly due to false detections in pastures. Grazing was also shown to be a confounding factor for mowing detection in other studies (Griffiths et al., 2020).

Previous studies have shown that mowing events can be detected through drops in vegetation index time series derived from optical images (Estel et al., 2018; Kolecka et al., 2018; Griffiths et al., 2020; Schwieder et al., 2022; Gómez Giménez et al., 2017). The main drawback of mowing detection with optical remote sensing is the irregularity of the time series in cloudy regions, causing omissions of mowing events.

In this chapter, we present a new method combining the completeness of S1 and the higher accuracy of S2 in a multi-source grassland mowing detection method. The method was developed as a module of an open-source toolbox, in the framework of the ESA-funded project Sentinels for Common Agricultural Policy (Sen4CAP), to facilitate the compliance assessment of several CAP subsidy schemes or support measures. The development and calibration of the method are not discussed extensively in this thesis.

The objective is to introduce the operational Sen4CAP grassland mowing detection method, evaluate its performance, and assess the potential of this multi-source approach for grassland monitoring over large areas. The method was first validated in the context of the Sen4CAP project in the six countries where it was developed. This validation was based on high-resolution image interpretation to rapidly build a large reference dataset. The method performances were then thoroughly assessed in the study area in Wallonia, using more precise

and complete reference data collected during the intensive field campaign. In addition, the grassland mowing detection by S1 was deeper investigated, more specifically regarding the impact of soil moisture on InSAR coherence and on the subsequent mowing detections, based on hourly precipitation data.

3.2 Method

The Sen4CAP mowing detection method consists of two independent object-based change detection algorithms based on S1 and S2 time series. Recent studies were considered as guidelines for feature and algorithm selection. The method was developed, calibrated, and refined in six pilot countries with varying climates and agricultural practices, namely Spain (ES), the Czech Republic (CZ), Italy (IT), Lithuania (LT), the Netherlands (NL), and Romania (RO). S2 mowing detection was tested with normalized difference vegetation index (NDVI), leaf area index (LAI), and fraction of absorbed photosynthetically active radiation time series (fAPAR). S1 mowing detection was tested with backscattering and coherence in both VV and VH polarization. The available mowing date information in the respective LPIS data sets of the pilot countries was used as a reference for the development, comparison, and calibration of the detection algorithms. The resulting optimal combination of S1 and S2 features was S2 NDVI and S1 VH coherence time series.

In this chapter, we describe and validate the operational grassland mowing detection method, available in the Sen4CAP toolbox v3.0 (Bontemps et al., 2020).

3.2.1 Satellite image processing

From S1, all IW L1C (Single Look Complex) images of the season, from S1 A and B satellites, in VH polarization, from ascending and descending passes intersecting the regions of interest, are downloaded. The SNAP S1-toolbox (ESA, v6.0) is used to process the images. The processing chain to obtain 6-day interferometric coherence is similar to the one described in section 2.3.2, except the averaging window size is 5 × 20 (in azimuth × range) and the final resampled and ground projected images have a spatial resolution of 20 m×20 m.

From S2, all available top-of-atmosphere L1C images of the season, intersecting the regions of interest, are downloaded. Atmospheric correction and cloud detection are performed using the MACCS ATCOR Joint Algorithm (MAJA). The multi-temporal cloud detection method implemented in MAJA allows to produce a more consistent and accurate cloud and shadow mask compared to the one obtained by the Sen2COR algorithm used by ESA (Hagolle et al., 2010; Baetens et al., 2019). The resulting L2A images and validity masks are then used to produce cloud-masked surface reflectance time series. Only images with less than 90% cloud cover are used. The NDVI is then calculated using the red (B4) and narrow NIR (B8A) bands (Equation 3.1).

$$NDVI = \frac{B8A - B4}{B8A + B4} \tag{3.1}$$

The obtained features are then averaged per parcel to build S1 coherence time series and S2 NDVI time series. A "no-touch" pixel sampling approach was applied, taking into account only pixels that are completely inside the parcel boundaries. This allows to limit border effects while guaranteeing a sufficient number of pixels per parcel to mitigate the speckle of the SAR imagery.

3.2.2 Mowing detection algorithms

The grassland mowing detection method is object-based and applied independently on each parcel. Two separate change detection algorithms are applied for S1 and S2 time series. Results are merged into a single output for each parcel, combining detections based on a reliability indicator.

Sentinel-1 algorithm

The interferometric coherence is expected to increase after a mowing event because of the shorter vegetation (cf. section **??**). Therefore, the S1 algorithm aims at detecting significant increases in coherence time series. Due to the growth of the vegetation, the coherence is expected to gradually decrease before a mowing event (Monti-Guarnieri et al., 2020; Morishita and Hanssen, 2014; Zalite et al., 2016). Therefore, each value $coh(t)$, extracted at time t, is compared to the value

$coh\_fit(t-1)$, predicted at time $t-1$ by linear fit of the six previous values $[coh(t-6),...,coh(t-1)]$. This allows to consider a potential slope in the coherence profile and detect sudden increases, compared to the previous signal trend. The detection is based on a Constant False Alarm Rate (CFAR) criterion (Equation 3.2).

$$coh(t) > coh\_fit(t-1) + k\sigma \tag{3.2}$$

The CFAR adaptive threshold ($k\sigma$) takes into account the standard deviation (σ) of the residual fitting errors, which are assumed to follow a Gaussian distribution in absence of a mowing event. The parameter k is fixed for a given probability of false alarm (PFA).

Sentinel-2 algorithm

The S2-based algorithm aims to detect significant decreases in NDVI time series while taking into account the irregular observation frequency due to cloud cover. Each observation $NDVI_{S2}(t)$, at time t, is compared with the last available cloud-free observation $NDVI_{S2}(t_{cf})$. The difference between $NDVI_{S2}(t)$ and $NDVI_{S2}(t_{cf})$ needs to be larger than a given threshold (Equation 3.3).

$$NDVI_{S2}(t) < NDVI_{S2}(t_{cf}) - th_{NDVI} \tag{3.3}$$

where th_{NDVI} is the fixed detection threshold.

To guarantee a minimum temporal precision despite the large gaps that can occur due to cloud cover, a maximum detection interval Δt_{max} is fixed. If an event is detected in the interval $[t_{cf}, t]$, the actual mowing is expected to have occurred no more than Δt_{max} days before the low NDVI value at time t. If the time $t - t_{cf}$ exceeds Δt_{max}, the detection interval in the output is then defined as $[t - \Delta t_{max}, t]$ instead. Table 3.1 shows the values of the detection thresholds and temporal parameters that were defined for the six pilot countries of the Sen4CAP project.

Figure 3.1 illustrates the detection of mowing events based on S2 NDVI and S1 VH coherence separately, on a grassland parcel in the Netherlands.

The confidence levels of S1 and S2 detections are estimated through a normalization function (Equation 3.4).

Table 3.1: Parameters for the S1 and S2 mowing detection algorithm and their calibrated values for the Northern (NL, LT, CZ) and Southern (IT, ES, RO) pilot countries of the Sen4CAP demonstration.

Parameter	Symbol	North	South
Probability of False Alarm (S1)	PFA	3.0×10^{-7}	
NDVI absolute decreasing threshold (S2)	th_{NDVI}	0.12	0.15
Minimum time between two detections	Δt_{min}	28 days	
Maximum detection interval (S2)	Δt_{max}	60 days	

$$f(x; min, max) = max - (max - min) \times e^{-x} \tag{3.4}$$

Where x is the difference $coh(t) - coh\_fit(t-1)$ and $NDVI_{S2}(t_{cf}) - NDVI_{S2}(t) - th_{NDVI}$ for S1 and S2 respectively. The limits $[min, max]$ are set to fit the confidence levels between separate range intervals ($[0, 0.5]$ for S1 and $[0.5, 1]$ for S2).

The confidence levels of overlapping S1 and S2 detections are then trivially merged. The four most confident detections, with a minimum interval Δt_{min} between $t_{start}(1)$ and $t_{end}(2)$ of consecutive detections, are retained. For each retained detection of each parcel, the time interval ($[t_{start}, t_{end}]$), the detection source (S1, S2, or S1+S2), and the confidence level are provided as output.

3.2.3 Study areas and reference data

This chapter presents the validation of the grassland mowing products obtained in the six pilot countries (ES, CZ, IT, LT, NL, RO) where the method was calibrated and a second independent validation based on the ground truth data collected in Wallonia (cf. section 2.3.1).

Planet image interpretation

The validation of the mowing detection method in the six pilot countries was based on Planet image interpretation to obtain dates of biomass removal on a large set of grassland parcels.

For each country, a sample of about 100 to 200 parcels was randomly selected from the national LPIS datasets to be visually interpreted. The random selection was stratified in order to be statistically representative of national grassland parcels distribution in terms

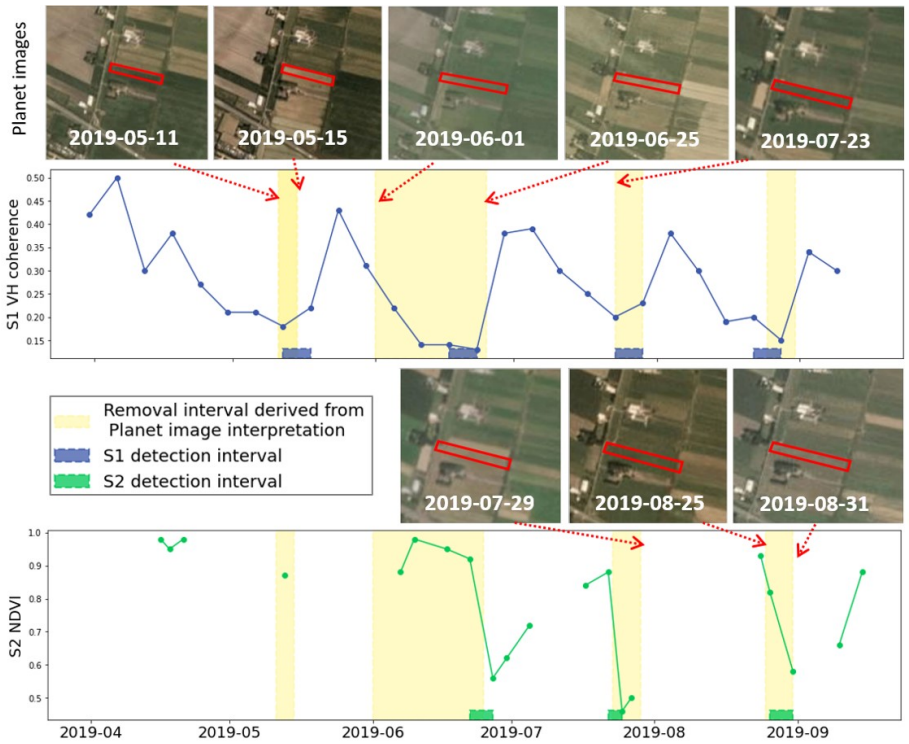

Figure 3.1: Detection of mowing events on a grassland parcel in the Netherlands, based on S1 VH coherence and S2 NDVI and illustration of the validation from Planet image interpretation.

of management (e.g. pastures and meadows, permanent and temporary) and vegetation type (e.g. Alfalfa, clover, presence of orchids) with a minimum of five parcels per class. An exception to this rule was made in Castilla y Leon (ES) because the most abundant class (82%) was 'grassland pastures', which are mainly managed by grazing and not mowing. The fraction of mown grassland parcels in the Spanish sample was increased to 40% to include enough mowing events for validation. The selected samples, forming a total set of 803 grassland parcels across the six countries, are described in Figure 3.2 and shown in Figure 3.3.

The reference mowing dates were then obtained through visual interpretation of daily true-color Planet images (average resolution: 3.5

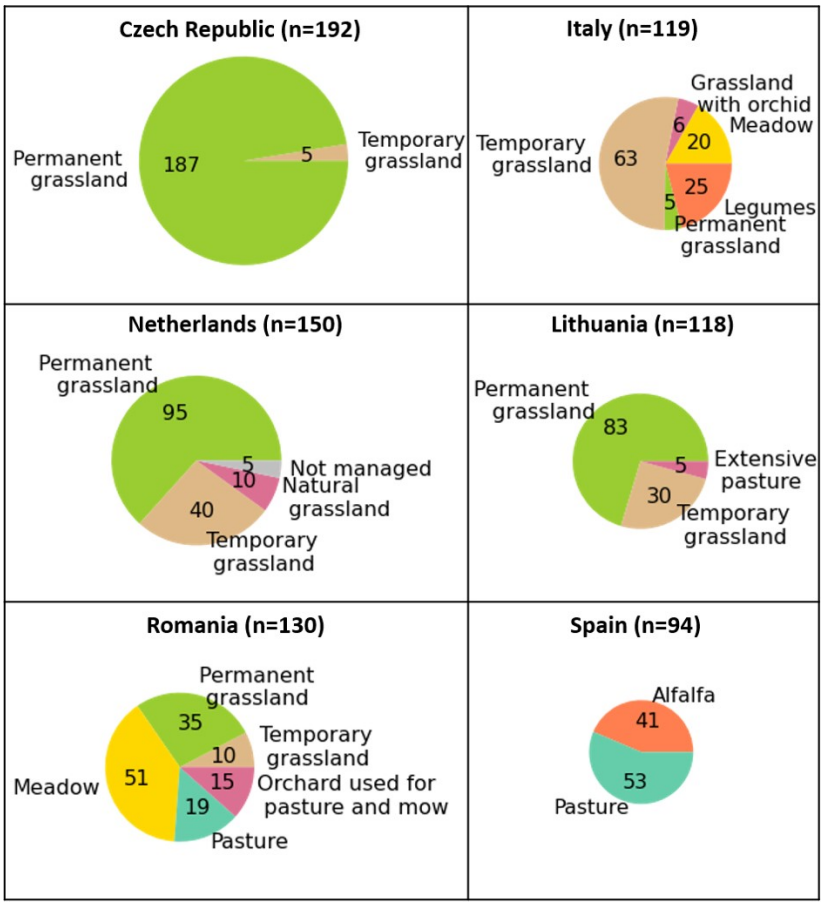

Figure 3.2: Distribution of parcels per grassland type selected for validation in each pilot country.

m). For each parcel, biomass removal intervals were identified, corresponding to the time interval between the last available cloud-free Planet image on which the grass seems to be tall (start date) and the first available cloud-free Planet image on which the grass seems to be short (end-date) (Figures 3.1 and 3.4 (a)). It is important to note that no clear distinction could be made between biomass removal by mechanical mowing or by intensive grazing.

Due to cloud cover, the length of the removal intervals varies, with a minimum of one day. A temporal buffer of three days was applied

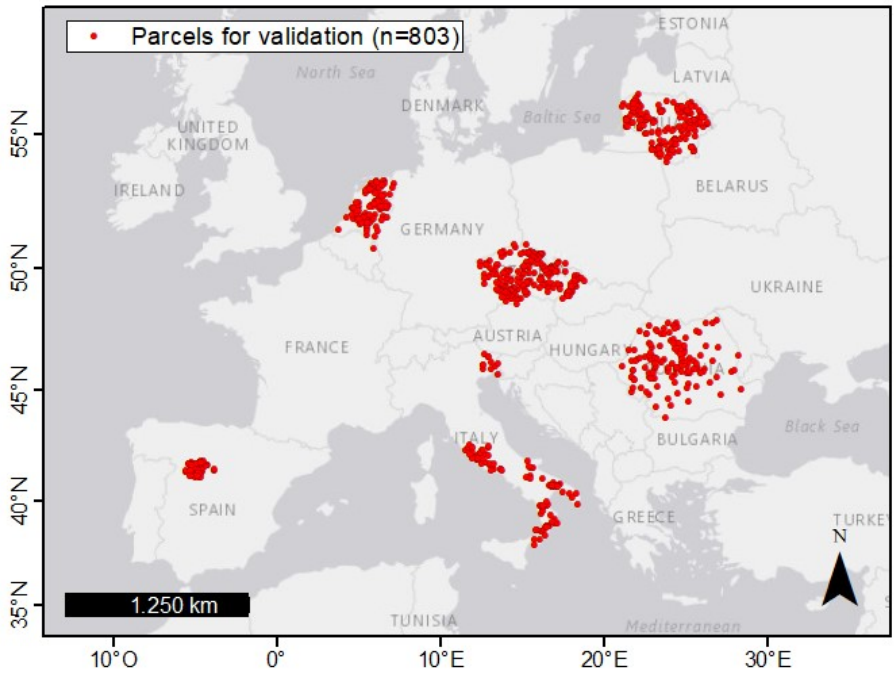

Figure 3.3: Overview of the geographical distribution of parcels selected in the six Sen4CAP pilot countries to validate the grassland mowing detection method with Planet image interpretation.

to the removal intervals to build the truth intervals for validation to compensate for the uncertainties inherent to the image interpretation. A longer buffer (10 days) was applied after the removal intervals to take into account potential delayed detections due to grass left on the field to dry (Figure 3.4 (a)).

Independent field campaign

The more precise and complete reference dataset collected during *in situ* the field campaign across Wallonia in 2019 (cf. section 2.3.1) was used for a fully independent performance assessment and to perform a sensitivity analysis of the Sen4CAP mowing detection method. This dataset includes a total of 4260 observation intervals (10 intervals for each of the 426 parcels) including grazing, mowing, and no activity intervals (Figure 3.4 (b-d)).

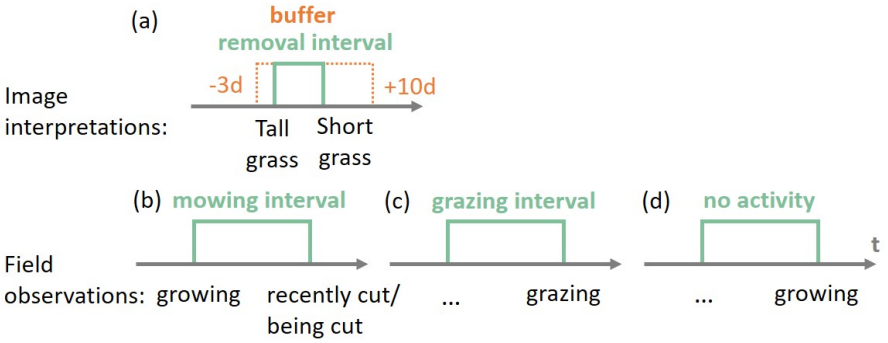

Figure 3.4: Schematic representation of the different truth intervals defined by Planet image interpretation (a) and from the observations made during the field campaign on 426 grassland parcels in Wallonia (b-d).

3.2.4 Validation

The performances of the mowing detection method were evaluated as in chapter 2, by crossing the reference datasets with detections at the temporal interval level to count true positives (TP), true negatives (TN) false positives (FP) and false negatives (FN) and compute the OA, the *detection rate* and the *precision* (Equations 2.10 and 2.11). As noted in the previous chapter, given the relative rarity of mowing events, the overall accuracy is dominated by the numerous TN and therefore yields an overoptimistic summary of the performance of mowing detection. In addition, the F1-score was computed as a combination of *detection rate* and the *precision* (Equation 3.5).

$$F1_{score} = 2 \times \frac{precision \times detection\,rate}{precision + detection\,rate} \tag{3.5}$$

3.2.5 Topsoil moisture as a potential confounding factor

In addition to the Sen4CAP mowing detection method validation, the impact of topsoil moisture on InSAR coherence was assessed as a potential confounding factor for the S1 mowing detection algorithm. This was done by estimating topsoil moisture at each S1 acquisition date and

performing statistical hypothesis and separability tests on coherence values obtained in different topsoil moisture classes.

The antecedent precipitation index (API) (Linsley et al., 1949) was used as a proxy for topsoil moisture. The API is a weighted summation of daily precipitation amounts. Hourly precipitation measurements were obtained from 11 weather stations of the Pameseb network of the Walloon Agronomic Research Center (CRA-W). Based on these measurements, the API at a time t was calculated as follows (Equation 3.6).

$$API_t = r \times API_{t-\Delta t} + P_{\Delta t} \tag{3.6}$$

where r is a recession coefficient representing the rates of drainage and evapotranspiration processes. The commonly used value $r = 0.84$ was applied (Zhao et al., 2019). $P_{\Delta t}$ is the cumulative precipitation over the time Δt, fixed here at one day.

Coherence time series were extracted for all permanent grassland parcels of the LPIS in a 5 km radius around each station, resulting in a total of 6966 parcels. The soil in the parcels around a station at time t was considered "wet" if API_t was in the upper quartile (0.75) of the station values and "dry" if it was in the lower quartile (0.25). Each six-day interferometric pair of dates at each station could thereby be characterized as "wet-wet", "dry-dry", "dry-wet" or "other" if it did not correspond to any of these three classes. The coherence values of each class were then compared through statistical hypothesis tests and separability measures. The tests were performed on S1 data from the descending and ascending pass separately, as the different acquisition times - respectively around 6 a.m. and 6 p.m. - could alter the impact of topsoil moisture on the signal (e.g. dew in the morning).

3.3 Results

3.3.1 Large extent validation based on Planet imagery

The main results of the validation in the six pilot countries of the Sen4CAP project are shown in Figure 3.5. These results show statistically significant differences in precision and detection rates among the countries. The differences could be related to the type of grassland parcels used for validation (cf. Figure 3.2) as well as to the climate and landscape characteristics of each country.

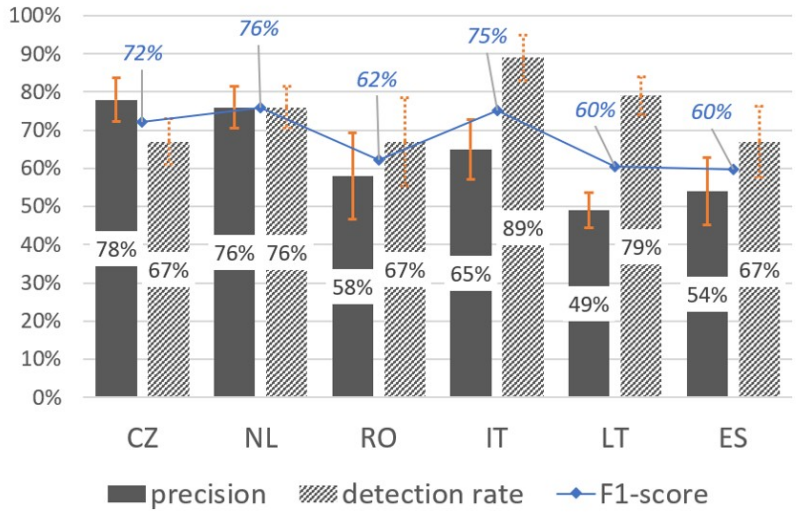

Figure 3.5: Precision, detection rate, and F1-score of the Sen4CAP mowing detection method estimated based on a visually interpreted reference dataset in six countries: Spain (ES), Czech Republic (CZ), Italy (IT), Lithuania (LT), the Netherlands (NL) and Romania (RO). The confidence interval of each metric is indicated in orange.

The best F1-score (76%) is achieved in the Netherlands, where the *precision* and the *detection rates* are equal. Italy and Lithuania have the largest *detection rate* (89% and 79%), but are penalized by a lower *precision*.

The highest *detection rates* occur in regions with a large number of temporary grasslands in the validation datasets. This can be explained by the larger biomass removal in the usually more productive temporary grasslands, but we did not have enough data to rigorously test this hypothesis in the scope of this study.

The highest *precision* is obtained in Czech Republic and the Netherlands. These two Northern countries also show the most balanced and confident performances. On the other hand, the prevalence of grazing practices could explain the lower precision obtained in Spain and Lithuania (54% and 49% respectively).

In Spain and Romania, the drying of grass in the summer also needs to be considered as a confounding factor for Planet image interpre-

tation, as well as for the mowing detection algorithms, especially if grasslands are managed through mixed grazing and mowing practices. In Italy, although present, drought appears to have less impact on the mowing detection performances. This can be explained by the prevalence of temporary grasslands, which are generally mown (not grazed) earlier in the season and therefore less affected by drought. The Southern countries (RO, ES, IT) are all characterized by a larger confidence interval, compared to the Northern countries (CZ, NL, LT). Additional analysis on the parcel size revealed no significant effect on the mowing detection performances across the pilot countries (Bontemps et al., 2021).

3.3.2 Field-based validation and performance assessment

In the 426 permanent grasslands observed during the field campaign in Wallonia, the Sen4CAP mowing detection method reached an overall accuracy of 97% and an F1-score of 58%. This independent validation showed a high *detection rate* of 83%, but the *precision* was relatively low, with only 44% of the detection intervals being true positives (i.e. actual mowing events).

A more detailed analysis of the false positives reveals potential sources of error (Table 3.2). First of all, almost half of the false positives overlapped a grazing interval in the reference dataset. In the remaining cases, no activity was observed at the time of the detection. When discarding all parcels where grazing was observed, and taking into account exclusively the mown parcels, *precision* increases to 73%, as a large part of false positives are removed. The *detection rate* is also slightly higher (85%). This results in an F1-score of 79% and an OA of 99%.

In terms of detection sources, 54% of the false positives were based on S1 only, the remaining being almost equally distributed between S2 and S1+S2 (24% and 22% respectively).

Finally, a majority (64%) of false positives occurred after the 20^{th} of June.

Further analysis of mowing detection sources shows the complementarity of S1 and S2. Figure 3.6 (a) shows the *precision* and *detection rate* obtained using S2 and S1 algorithms alone, compared to the results achieved by the combined algorithms. The *precision* is highest

Table 3.2: Detailed analysis of false positives, highlighting potential sources of error for the Sen4CAP mowing detection using S1 and S2. Number (n) and fraction (%) of false positives per (i) observed activity at the time of detection, (ii) detection source of the detection, and (iii) detection before or after June 20.

FALSE POSITIVES (n=268, commission error = 56%)		
Field observation	n	%
no activity	139	52%
grazing	129	48%
Detection source		
S1	145	54%
S2	64	24%
S1+S2	59	22%
Timing of the detection		
early event (≤ June 20)	91	34%
late event (> June 20)	177	66%

when using S2 alone (59%), compared to 45% with S1 alone and 44% with both. The *detection rate*, however, is largely improved, up to 83% with combined detections, compared to 69% for each algorithm. Figure 3.6 (b) shows the distribution of true and false positives per detection source when applying the combined algorithms. 80% of detections by S1 alone are false positives. The number of detections by S2 alone is relatively low, with a small majority of false positives. The most certain detections are those confirmed by both satellites (72% TP for S1 + S2). Most true positives were detected by S1 and S2 (64%), while 17% were detected by S1 alone, and 19% by S2 alone. These balanced results highlight the complementarity of the two satellites.

The confidence level computed for each detection provides valuable information about their certainty. Figure 3.7 shows the *precision* and the number of detections for different confidence level intervals. Most detections have a confidence level between 0.4 and 0.6 and between 0.8 and 0.9. Under a confidence level of 0.4, less than 10% of the detections are true positives. Above a confidence level of 0.2, the *precision* is strongly correlated with the confidence level ($R^2 = 0.94$).

Finally, the mowing detection performances were also analyzed per parcel to assess the ability of the method to identify different types of

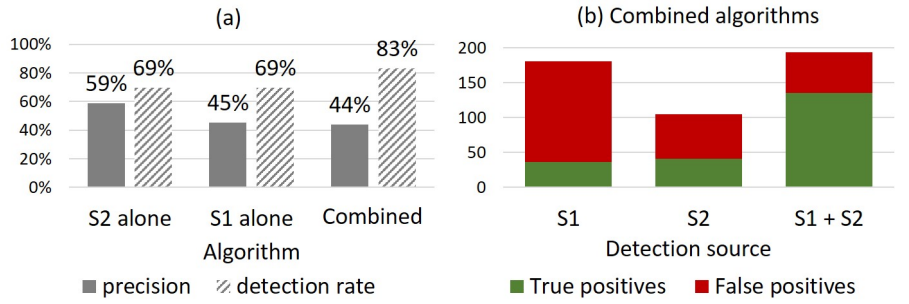

Figure 3.6: (a) Mowing detection performances (*precision* and *detection rate*) of algorithms based on S1 and S2 time series alone compared to the combined approach. (b) Distribution of true and false positives per detection source (S1, S2, or S1+S2) when applying the merged approach.

mowing practice observed during the field campaign in terms of frequency and precocity of mowing events (ME) ("no ME", "1 early ME", "1 late ME" and "2 ME"). Figure 3.8 (a) shows the results considering all detections. As expected based on previous analyses, most false positives occur in parcels that were not mown during the study period, resulting in a low fraction (32%) of correctly identified "no ME" parcels. From the parcels with 1 early ME and 1 late ME, respectively 58% and 76% are identified correctly. Most errors in these classes are due to an additional false positive. In the "2 ME" class, 76% of the parcels are identified correctly and almost all errors are due to the omission of the first (early) ME.

The same evaluation was performed after filtering out detections with confidence levels below 0.5 to reduce the number of false positives (Figure 3.8 (b)). This significantly improves the results on "no ME" parcels, as more than twice as many are then identified correctly (69%). Although more ME are omitted, performances are also slightly improved for parcels with 1 ME (from 58% to 60% for early ME and from 78% to 83% for late ME). Performances on parcels with 2 ME, however, are much lower (24%) as the first ME is often omitted. Given that most omissions appear to occur for early ME, which are crucial for grassland use intensity assessment, the minimum confidence level could be adapted to the time of the season.

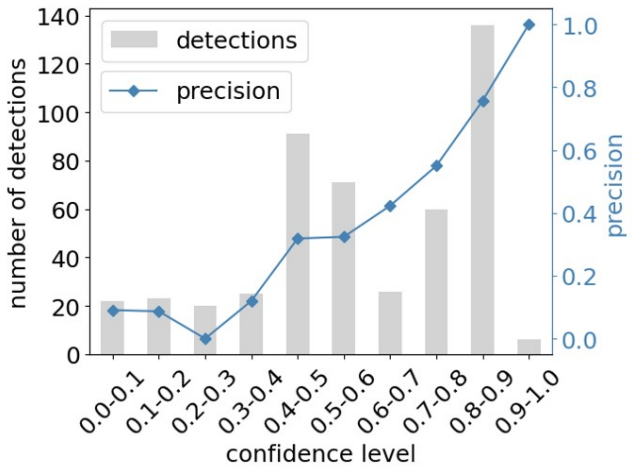

Figure 3.7: *Precision* and number of detections per given confidence level.

A third test was performed considering early detections with a minimum confidence level of 0.4 and late detections with a minimum of 0.5 (Figure 3.8 (c)). This improves the results in parcels without ME (67%) while maintaining a reasonable *detection rate* of early events (74%) in parcels with "2 ME". With this adaptive minimum confidence level, 69% of all parcels are identified correctly in terms of mowing practice, compared to 51% when taking all detections into account.

3.3.3 Algorithm sensitivity

In order to assess the sensitivity of the algorithm, it was tested with different parameter configurations in the Walloon study area. The parameter values defined in Table 3.1 for the Northern countries were used as a baseline. We compared the *detection rates* and *precisions* obtained when increasing or decreasing these parameter values.

The impact of different values of PFA and th_{NDVI} on *detection rate* and *precision* is shown in Figures 3.9 (a) and (b). When increasing the PFA for S1 mowing detection from 3×10^{-7} to 3×10^{-6} or 3×10^{-5}, the loss in *precision* is greater than the gain in *detection rate*, reducing the overall performance. When decreasing the PFA to 3×10^{-8}, the *precision* is slightly enhanced, while the *detection rate* is a bit lower. When

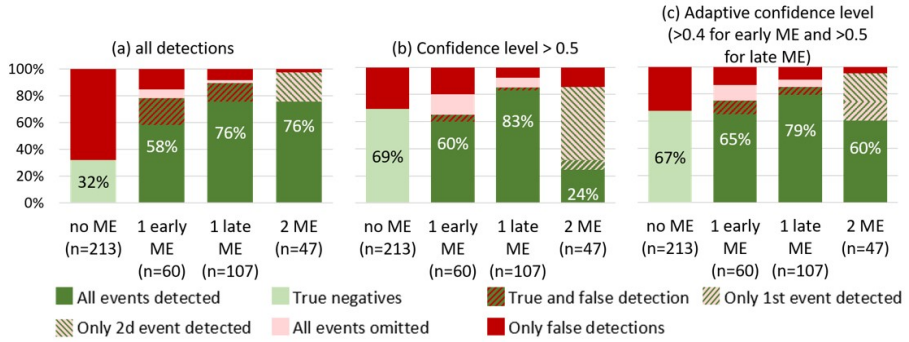

Figure 3.8: Mowing detection evaluation per type of observed mowing practice, considering (a) all detections, (b) detections associated with a confidence level above 0.5, and (c) using an adaptive minimum confidence level for early and late detections. For each option, the fraction of correctly estimated parcels is given, along with the types of errors that occur (false positives, omissions, or both).

changing the th_{NDVI} for S2 detection to 0.10, 0.14 or 0.20, the loss is systematically more important than the gain in *precision* or *detection rate*. Overall, the changes in performances are relatively small in the range of tested threshold values, showing a small sensitivity to the algorithm parameter values.

The analysis of the temporal parameters Δt_{max} and Δt_{min} is shown in Figure 3.10. With a maximum detection interval of 60 days, most detection intervals don't exceed 24 days (Figure 3.10 (a)), although the region is relatively cloudy, meaning a Δt_{max} of 60 days may be excessively cautious. When Δt_{max} is decreased to 18 days and thus reaches a higher temporal precision, there is no loss in *detection rate* (Figure 3.10 (b)). However, at a Δt_{max} of 12 days, some detections become false positives, since the true mowing event was observed earlier. This results in a lower *detection rate* and lower *precision*. On the other hand, the minimum time Δt_{min} of 28 days between two consecutive detections appears to be an adequate choice (Figure 3.10 (c)). Reducing Δt_{min} (to 21 or 14 days) or increasing it (to 35 days) has an overall negative effect on the detection accuracy.

These different results confirm that the standard set of parameter values (Table 3.1) for the Northern European countries seems the most appropriate.

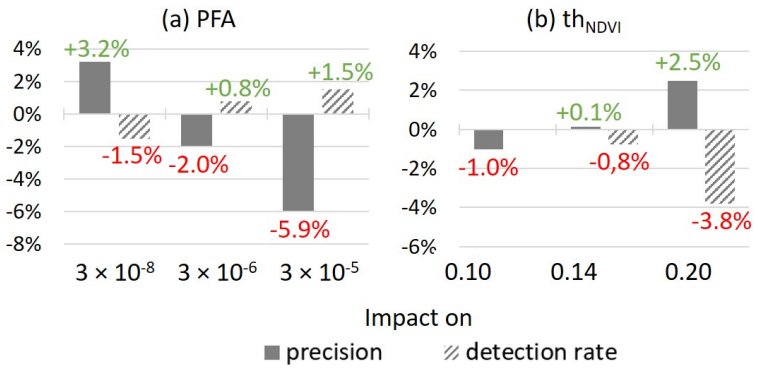

Figure 3.9: Sensitivity of the mowing detection algorithm calibration. Impact of two threshold parameters: (a) probability of false alarm (PFA) for S1 detection and (b) NDVI decreasing absolute threshold (th_{NDVI}) on *detection rate* and *precision*, compared to the calibrated parameter values.

3.3.4 Impact of soil moisture on interferometric coherence

The distributions of S1 coherence values of grassland parcels in different topsoil moisture classes ("wet-wet", "dry-dry", "dry-wet" and "others") were compared to assess the impact of soil moisture on coherence (Figure 3.11). Although the distributions of the four classes have significant overlaps, coherence values are slightly higher on average for the "dry-dry" class than for the "wet-wet" or "dry-wet" classes. Coherence is also slightly higher for the "wet-wet" class than for the "dry-wet" one.

The coherence distributions of each class were compared through a statistical hypothesis test. The resulting p-values are given in Table 3.3, along with the histogram intersections and the absolute differences between the averages (bias). The p-values are extremely low because of the large sample sizes ($n_{min} \approx 11000$) used to compute the averages, which are indeed different from one class to another. The histogram intersections are however high, showing very limited separability between the coherence values of the different topsoil moisture classes. The smallest intersection (71%) is between "dry-dry" and "dry-wet". These two classes also show the largest bias between their respective average coherence values, with coherences higher by 0.06 for the "dry-dry" class. As it is more diverse, the "others" class shows the least dif-

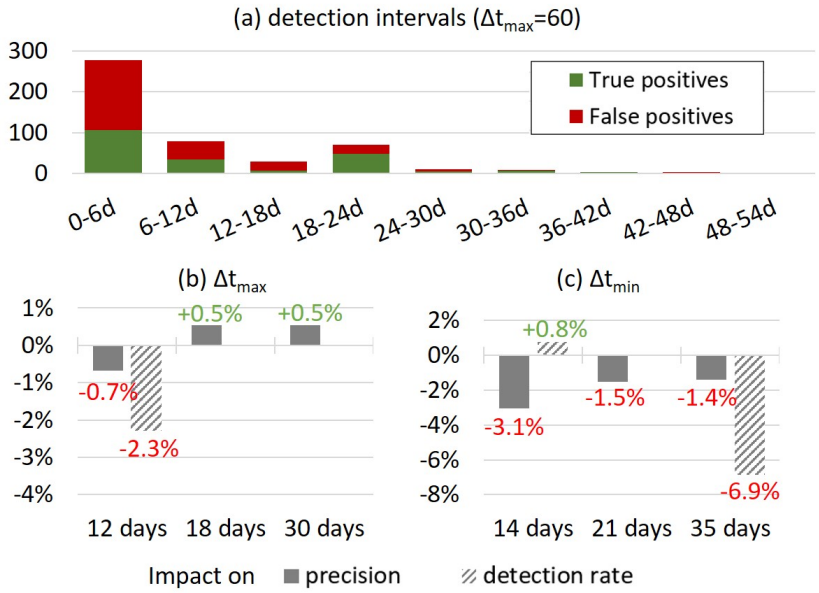

Figure 3.10: (a) Distribution of true and false positives per detection interval category Δt. Impact of (b) maximum detection interval (Δt_{max}) and of (c) minimum interval (Δt_{min}) between consecutive detections on *detection rate* and *precision*, compared to the calibrated parameter values.

ference and separability from the three extreme classes. Overall, these results show that slightly higher coherence values can be expected on grasslands during consecutive dry periods. Important precipitations, resulting in "dry-wet" and "wet-wet" classes, cause slightly lower coherence on average, due to signal decorrelation.

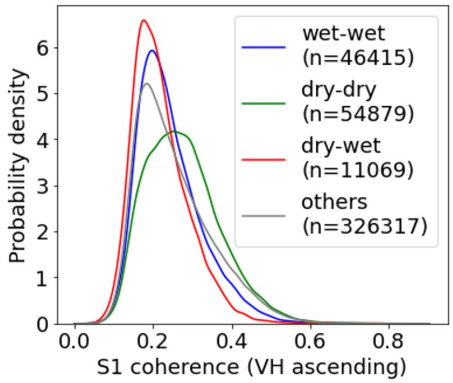

Figure 3.11: Distribution of S1 coherence (VH ascending) values for different topsoil moisture classes in grassland parcels.

Table 3.3: Comparison of S1 coherence (VH ascending) values distributions in different soil moisture classes. In each cell, the table shows the statistical T-test p-values (top), the histogram intersection (center), and the bias (bottom) between classes.

COHE	dry-dry	dry-wet	others
wet-wet	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$
	82%	81%	91%
	-0.03	0.02	-0.01
dry-dry		$< 10^{-5}$	$< 10^{-5}$
		71%	86%
		-0.06	0.02
dry-wet			$< 10^{-5}$
			84%
			-0.03

3.4 Discussion

3.4.1 Complementarity of S1 and S2

Optical vegetation index time series have been used previously for detecting mowing events, providing overall satisfying results (Griffiths et al., 2020; Kolecka et al., 2018; Schwieder et al., 2022). These methods, however, rely on the availability of cloud-free images before and, more importantly, after an event. Depending on the region of interest, this condition cannot always be met. In the study area in Bel-

gium, the average monthly cloud frequency (based on MODIS 1 km resolution images from 2000 to 2014) in May, June and July, is as high as 73%, 75%, and 70% of cloudy days, respectively (Wilson and Jetz, 2016). Such a persistent cover, which is common in northern European countries, strongly reduces the actual observation frequency of S2. Griffiths et al. (2020) used the Harmonized Landsat Sentinel-2 (HLS) dataset (from 2016, before Sentinel-2B was launched) and compositing to build a regular NDVI time series and obtained promising results. Their biggest drawback was, however, the significant number of omitted mowing events. Schwieder et al. (2022) developed a similar method, based on HLS (with the complete Sentinel-2 constellation) and detecting anomalies, compared to an idealized grassland phenology curve. Their results on hay meadows (F1-score 58-67%) are comparable to those obtained in this study, on hay meadows and pastures, when using S2 alone. Their detection method however included more conservative detection rules and thresholds to avoid false detections (e.g. due to unmasked clouds), gaining in *precision*, but omitting more events. Shorter revisit cycles (e.g. daily acquisitions) would improve detection performances to some extent. However, with 21 to 23 cloudy days per month, the probability of having extended periods without a clear image — and thereby omitting mowing events — remains high.

Therefore, SAR imagery represents a great asset for mowing detection in cloudy areas, as microwaves are transmitted through clouds. Our results are consistent with previous studies, showing that mowing events can be detected through jumps in InSAR coherence time series (De Vroey et al., 2021; Voormansik et al., 2020; Tamm et al., 2016). Detections by S1 are, however, less precise than those by S2. This was confirmed by our analysis of the detection sources in the Walloon study area. The SAR signal can be impacted by many factors outside the reduction of biomass, such as soil and vegetation moisture. On the descending pass, images are acquired in the early morning (around 6 a.m.). Since both passes are used to build S1 time series for detection, the effect of morning dew on the SAR signal could be a possible explanation for the lower performance. The signal decorrelation due to increased soil moisture, thus preventing coherence to rise after a mowing event, is a potential cause for omissions (section 3.3.4). Applying a systematic bias correction would, however, require a cautious investigation because of the variability of the soil moisture impact. Overall,

the complex interaction of SAR signal with objects and surfaces, and the inherent speckle of SAR imagery, make it more challenging to interpret.

Despite the challenges inherent to SAR-based methods, the combination of S1 and S2 detection algorithms in this multi-source method significantly increased the *detection rate*. Some of the events that were omitted by S2 due to cloud cover could be detected by S1. On the contrary, events that were omitted by S1 due to a change in soil moisture or other confounding factors could be detected by S2 when a clear image was available shortly after the event.

Moreover, the estimated confidence levels, based on the normalization of the amplitude of the signal change and used to merge the S1 and S2 detections, were shown to be well correlated to the *precision* of the detection. They are consequently a good indicator of the probability of occurrence of a mowing event and can be used to screen out detections according to the final use, to be more or less conservative.

The complementarity of optical and radar imagery for mowing detection was also recently confirmed by Lobert et al. (2021) with a deep-learning approach. In their study, they compared different combinations of S1, S2, and Landsat-8 features in a CNN. All combinations of optical and SAR features outperformed exclusive uses of optical or SAR features. NDVI and coherence were their fifth best performing combination, with a detection rate (recall) of 85% and precision of 79%. From a fully independent validation, we obtained comparable results on hay meadows in Wallonia, using similar input features in a versatile and transparent rule-based change detection method.

Overall, recent studies on automated grassland mowing detection seem to show coherent results and conclusions. Multi-source mowing detection methods combining optical and SAR data should be further investigated and developed.

3.4.2 Reference data and validation

Although previous studies showed the feasibility of mowing detection through satellite remote sensing, they often lacked sufficient precise and complete reference data to calibrate and validate their methods (Reinermann et al., 2020). Information on mowing dates and practices

is rarely readily available and collecting it in situ is extremely time-consuming.

Image time series interpretation offers a cost-effective and less time-consuming alternative for grassland mowing detection validation, as a unique reference dataset or as a complement to smaller samples of reported or observed mowing events (Griffiths et al., 2020; Schwieder et al., 2022). In this study, the Planet image interpretation approach allowed to rapidly gather a large reference dataset (n=803) to validate the mowing detections in six countries throughout the whole season (April to October 2019). However, the reliability of this reference dataset was limited for several reasons. First, persistent cloud coverage prevented the observation of some events in northern countries. Secondly, the absence of the NIR band in the Planet Web Mapping Service available for the Sen4CAP partners limited the certainty of observations due to confusion with ploughing events and grassland droughts in the Southern countries. Finally, the varying radiometric and geometric precision of Planet images, depending on the satellites that acquire the data, increased the uncertainty of the interpretation.

Overall, there is a risk of non-independence between the reference data and the classification results due to the similarity of criteria used by the machine and by the image interpreters. This non-independence could introduce a bias in the estimation of accuracy indices (Radoux and Bogaert, 2020).

The second validation dataset acquired through an intensive field campaign is much more complete, more accurate, and fully independent of the classification rationale. This dataset allowed to estimate the *detection rate* and the *precision* of the mowing detection method with more certainty, to test its transferability and evaluate its potential for grassland use intensity assessment. While the results of the Planet-based validation reflect the ability of the method to detect grassland biomass removal, regardless if by mowing or intensive grazing, the field dataset of Wallonia allowed to validate the detection of actual mowing events (differentiating them from grazing). During the field-based validation, mowing detections overlapping grazing intervals were considered as false positives, which strongly reduced the computed *precision*. Half of the false positives were confirmed as intensive grazing events. Such detections would be considered as true

positives in the Planet-based validation since there was no distinction between mowing and grazing in the reference dataset.

In addition to — and depending on — the source and quality of reference data, the performances of a mowing detection method can be evaluated at different levels. In this study, we used a conservative validation approach, which reflects the ability to exactly detect the occurrence and the timing of each mowing event and allows to compute explicit metrics such as the *precision*, *detection rate* and F1-score (Lobert et al., 2021; De Vroey et al., 2021; Schwieder et al., 2022). With such an approach, a delayed or untimely detection is counted as a false negative and a false positive. Therefore, we performed an additional analysis per parcel to assess the ability of the method to identify different types of mowing practice in terms of frequency and precocity of ME. Another possible approach is to validate the frequency and exact timing of mowing events separately (Griffiths et al., 2020; Schwieder et al., 2022). With this approach, a delayed (or untimely) detection is only considered an error in terms of timing, since the mowing frequency is accurate. In general, a combination of different levels of performance evaluation provides the best and most complete estimation of a method's potential and limitations and allows a comparison of results between studies.

3.4.3 Diversity of practices

Given the diversity of climates, landscapes, and grassland management practices across Europe, it is challenging to develop a grassland mowing detection method that is adapted to all regions. Part of the dissimilarities in mowing detection performances between the pilot countries can indeed be explained by climate. While more mowing events are omitted in Northern regions due to cloud cover, drought can cause false detections and reduce the precision of the method in Southern regions.

The method accurately detects the removal of grass biomass and works best on permanent grasslands managed as hay meadows. Most studies on grassland mowing detection focus on permanent hay meadows. However, from an operational point of view, we wanted to evaluate its performance in different types of grasslands because precise information on land cover and land use is not always available. The

various types of grasslands in the pilot countries, i.e. more or less intensive, temporary or permanent, pastures or meadows, vegetation types, etc. can also explain part of the discrepancies in the results.

The performances of the mowing detection method in different grassland types could however not be assessed quantitatively, because of the limited available information and the lack of uniformity between the countries' respective LPIS grassland categories. In many cases, the categories remain subject to interpretation in terms of management practice and intensity. For example, in Romania, a distinction is made between meadows and pastures, while some grasslands are classified more generally as permanent grasslands. In Italy and Spain, grasslands include leguminous crops. In the Czech Republic, the Netherlands, and Lithuania, a majority of parcels are classified as permanent grasslands, which could be either mown or grazed, more or less intensively.

The field dataset collected in Wallonia was restricted in terms of land cover to permanent grasslands dominated by gramineous plants but included all types of agricultural land use, namely pastures, meadows, and mixed practices. The regular field observations provided more precise information about the management practices of each parcel. This allowed to analyze the mowing detection performances with regard to management practices. In any case, the SEN4CAP algorithm performed as expected. On one hand, under the same conditions, late ME are better detected than early ME. Detection may be hindered because of the faster regrowth of grass in the spring, making it more challenging to spot the smaller decrease in biomass. Detecting early mowing events is, however, crucial for grassland use intensity assessment. Increasing the *detection rate* to capture the early ME however implies to sacrifice some *precision*. On the other hand, grazing practices that engender a large biomass removal in a short period (e.g. intensive rotational grazing) are a major confounding factor. Without information about the location of hay meadows, increasing the *precision* in areas with intensive pasture, therefore, requires sacrificing the *detection rate*.

Other practices can potentially prevent the accurate detection of mowing events by S1 and S2. For example, when cut grass is left on the parcel to dry for a few days after a mowing event, it can influence the signal and prevent an accurate change detection.

Finally, it must also be considered that the management practices might not be homogeneous at the parcel level. Indeed, the full par-

cel, as declared in the LPIS, is not always mown at the same time (e.g. in the case of rotating intra-parcel management). In an object-based approach, partial mowing of parcels risk being omitted as both mown and unmown pixels contribute to the average parcel value, reducing the apparent change in the time series. Pixel-based mowing detection approaches using optical imagery have shown promising results (Griffiths et al., 2020; Kolecka et al., 2018; Schwieder et al., 2022). Monitoring grasslands at the pixel level allows to account for intra-parcel variability in terms of practices, but on the other hand, it causes a salt-and-pepper effect that is avoided in an object-based approach. Furthermore, while optical pixel-based mowing detection is feasible, the speckle effect inherent to SAR imagery would probably be a significant issue for SAR pixel-based mowing detection. Furthermore, the combination of S1 and S2 mowing detection at the pixel level would be more complex and would imply a resampling step. Another option would be to use S1 and S2 data fusion methods simulating optical images (Garioud et al., 2021; He and Yokoya, 2018) to work at the pixel level while compensating for cloud cover.

Overall, the multi-source mowing detection method presented here detects mowing events with relatively high accuracy and allows to consistently differentiate various mowing practices and grassland use intensity trends across different agroecological regions. It can be adapted by changing the detection threshold parameters to optimize the balance between *detection rate* and *precision*, depending on the context and objective of its use. Furthermore, the final result can be filtered, based on the provided confidence level, as it is a good indicator of the reliability of the detections. The mowing detection method of Sen4CAP could thereby be used for large-scale grassland monitoring and even mapping ecological habitat quality.

3.5 Conclusion

This study showed the full potential and limitations of Sen4CAP's multi-source mowing detection method based on S1 and S2 time series. The exhaustive reference datasets allowed to show the consistency of the algorithm across seven European countries and various types of grasslands, while highlighting the importance of reference data quality.

Thanks to the complementary use of S1 and S2, the method reached a *detection rate* of 85% and a *precision* of 73% (F1-score 79%) for detecting mowing events on hay meadows. Furthermore, the detection of mowing events along the growing season allows to classify mowing practices with an overall accuracy of 69% and should allow to differentiate grasslands in terms of management intensity. Further efforts are still needed to improve the accuracy of mowing detection and produce a more precise and thematically complete grassland classification. Nevertheless, this adaptive and transparent mowing detection method could be implemented in large-scale grassland monitoring. Combined with ecological modeling tools, it could be used to support agro-environmental schemes in Europe.

Chapter 4

Classification of grassland management units using optical and microwave remote sensing

This chapter is adapted from the following article:

De Vroey, M., Radoux, J. & Defourny, P. (2023). Classification of grassland management units using optical and microwave remote sensing *Remote Sensing*, 15, 181. 10.3390/rs15010181

Abstract

Most European temperate grasslands are managed through mowing, grazing, or a combination of both in relatively small management units. Grazing and mowing can however not be considered equivalent, since the first is gradual and selective and the second is not. In this chapter, the aim is to differentiate grasslands in terms of management practices and to retrieve homogeneous management units. Grasslands are classified hierarchically, first through a pixel-based supervised classification to differentiate grazed pastures from mown meadows and then through an object-based mowing detection method to estimate the management intensity. A large ground truth dataset was used to calibrate and validate the method. A strict mask was created to only consider pure grassland pixels without shadow. The best classification was obtained with a random forest classifier based on interpolated LAI times series, differentiating pastures (grazed) from hay meadows (mown) with an overall accuracy of 88%. The classification is then combined with existing parcel delineation to retrieve homogeneous management units, which are used for object-based mowing detection based on Sentinel-1 coherence and Sentinel-2 NDVI. The mowing detection performances were increased thanks to the grassland mask, the management unit delineation, and the exclusion of pastures, reaching a precision of 93% and a detection rate of 82%. This hierarchical grassland classification approach allowed to differentiate 3 types of grasslands, namely pastures, and meadows (including mixed practices) with an early first mowing event and with a late first mowing event, with an overall accuracy of 79%. Grasslands could be further differentiated by mowing frequency resulting in 5 final classes.

4.1 Introduction

Most grasslands in temperate areas are managed through grazing, mechanical mowing, or a combination of both for forage production. The stocking rate of grazing animals and the timing and frequency of mowing events are major factors of grassland use intensity. The type and intensity of management practices are commonly used as indicators to classify grasslands, both from an agricultural and an ecological perspective (Dufrêne and Delescaille, 2005; Blüthgen et al., 2012; Tonn et al., 2020).

In previous chapters, the focus was set on automated mowing detection methods. These methods and most other mowing detection methods described in the literature are object-based, relying on existing delineations of parcel boundaries or a preliminary segmentation step (Tamm et al., 2016; Gómez Giménez et al., 2017; Voormansik et al., 2020; De Vroey et al., 2021; Lobert et al., 2021; De Vroey et al., 2022). This allows a spatial smoothing of the satellite signal and avoids the salt and pepper effect that is inherent to pixel-based approaches (Kolecka et al., 2018; Griffiths et al., 2020; Schwieder et al., 2022). One of the major drawbacks of object-based approaches is however the potential heterogeneity of practices inside declared or delineated parcels. When only one part of a parcel is mown at a given time and the other is grazed or mown at a different time, the signal change can be smoothed out, causing omissions.

While many studies focused on mowing detection, few have investigated the possibility of monitoring grazing activities (Franke et al., 2012; Gómez Giménez et al., 2017; Zheng et al., 2018) and even fewer discriminate grazed and mown grasslands (Stumpf et al., 2020). Grazing has been identified as a major confounding factor to mowing detection in several studies, as many false mowing detections occur in pastures (De Vroey et al., 2021; Griffiths et al., 2020) because they both result in biomass removal. However, both from an agricultural and an ecological point of view, grazing and mowing cannot be considered equivalent management practices, since the first is selective (depending on the type of livestock) and usually gradual (except for intensive rotational pasture systems), while the second is not. Moreover, on pastures, the effects of e.g. trampling and manure input need to be considered.

The objective of this study is to differentiate grasslands in terms of management practices and to retrieve homogeneous management units. Grasslands are differentiated in two hierarchical steps, combining a pixel-based and an object-based method to account for the variability of practices inside declared parcels.

The first step consists in differentiating two types of managed grasslands at the pixel level, namely (a) pastures, managed exclusively through grazing, and (b) hay meadows, which are mown mechanically and sometimes also grazed after a first mowing event (i.e. mixed practices). This preliminary pixel-based classification approach tackles two main issues in grassland monitoring; (i) the heterogeneity of practices inside declared parcels and (ii) grazing as a confounding factor for mowing detection.

Then the object-based mowing detection method based on Sentinel-1 and Sentinel-2 time series (Chapter 3) is applied to the management units classified as hay meadows to further differentiate them and to produce an exhaustive grassland management practice classification.

4.2 Materials

4.2.1 Satellite data

The grassland classification method developed in this chapter is based on Sentinel-1 SAR and Sentinel-2 multi-spectral optical time series.

During the study period (April 9 to July 19, 2019), 17 scenes were acquired by Sentinel-1 A and B in ascending pass and 18 in descending pass over the study area (Table 4.1). All scenes were processed from Single Look Complex to Ground Range Detected γ_0 backscattering amplitudes, using SNAP Sentinel-1 toolbox (ESA, v6.0). The processing chain is shown in Figure 2.4.

Sentinel-2 A and B acquired 17 multi-spectral images with less than 80% cloud cover over the study area (tile 31UFR) during the study period (April 9 to July 19, 2019). The top of atmosphere images (Level 1) were converted to surface reflectances (Level 2) using Sen2Cor v2.10 for atmospheric correction. The Function of mask (Fmask) method (Zhu and Woodcock, 2012) was used for cloud masking since it was shown to be more accurate than the Sen2Cor cloud masking method (Baetens et al., 2019).

4.2.2 Field data

The reference data for training and validation was obtained from the field campaign carried out across Wallonia in 2019 (Section 2.3.1). Based on the observations, four classes of grasslands could be differentiated: parcels where only grazing was observed (pastures, n=201), grazed parcels with at least one mowing event (mixed practices, n=61) and parcels with no grazing, but at least one mowing event before (hay meadows ($<$ 15/06), n=78) or after June 15th (extensive hay meadows ($\geq$ 15/06), n=76).

4.3 Method

Previous chapters have shown the great potential of combining Sentinel-1 and Sentinel-2 for grassland mowing detection. Because of the speckle inherent to SAR imagery, pixel-based approaches are challenging. Therefore methods using SAR data for mowing detection rely on object-based approaches, averaging the signal per parcel.

However, grassland parcels, as declared by farmers in the LPIS often include several management units that are not all exploited at the same time or in the same way. To illustrate, figure 4.1 shows the Sentinel-2 derived Leaf Area Index (LAI) time series of a grassland parcel in our study area. The parcel's average time series (in gray) is relatively constant, while about half of the pixel time series (in green) significantly decreases in the middle of June. This decrease in LAI is due to a mowing event that occurred on one part of the declared parcel, while the other part (in yellow) was not mown, but grazed. The two management units are also visible on the orthophoto (dashed red line on figure 4.1).

This is an issue for object-based grassland monitoring methods such as the one developed in Sen4CAP (Chapter 3). Therefore, we develop a hierarchical grassland characterization approach combining a pixel-based classification method and an object-based mowing detection method.

4.3.1 Pixel-based supervised classification

In this phase, the aim is to differentiate grasslands in terms of main management practice (grazing or mowing) and retrieve homogeneously

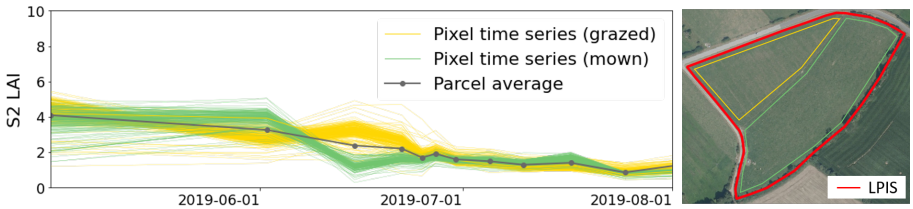

Figure 4.1: Leaf Area Index (LAI) time series (retrieved from Sentinel-2) extracted per pixel and per parcel (average value) for a grassland parcel (LPIS delineation in red on the orthophoto (SPW)). About half of the parcel was mown (in green), while the other part was grazed (in yellow).

managed parcels for further characterization. A pixel-based supervised classification method is used to discriminate exclusively grazed *pastures* from mown *hay meadows*. The *hay meadows* class includes mixed practices, which are alternatively mown and grazed. The field observations are used to build a reference dataset to train and validate a random forest classifier based on S2 and S1 time series.

Input features

Vegetation indices derived from specific spectral bands are commonly used to emphasize certain properties of a vegetation cover, such as biomass. We made the hypothesis that grasslands that are grazed throughout the season should have relatively constant and stable vegetation index time series compared to grasslands with at least one mowing event causing a sudden change in biomass (Griffiths et al., 2020; Schwieder et al., 2022). To test the sensitivity of the classification to the choice of vegetation index, three spectral vegetation indices and one biophysical index derived from Sentinel-2 were considered for the classification, namely NDVI, the red edge chlorophyll index (CIre), and the leaf area index (LAI). The NDVI is computed as the normalized difference between the near-infrared (band 8) and the red (band 4) reflectance (Equation 4.1) and is largely used for vegetation monitoring and more specifically for grassland mapping and mowing detection (Estel et al., 2018; Griffiths et al., 2020; Kolecka et al., 2018). The CIre is related to the increase in reflectance between the red and near infra-red

(i.e. the red edge) which is linked to biomass and chlorophyll content of vegetation. It is calculated as the ratio between lower (band 5) and upper (band 7) red edge reflectance (Equation 4.2). The CIre has been used by Hardy et al. (2021) to retrieve grassland biomass.

$$NDVI = \frac{Band\,8 - Band\,4}{Band\,8 + Band\,4} \tag{4.1}$$

$$CIre = \frac{Band\,7}{Band\,5} - 1 \tag{4.2}$$

The LAI was retrieved from Sentinel-2 reflectances through the calibrated artificial neural network from the BV-NET tool (Weiss and Baret, 1999), which is implemented in several ESA agricultural monitoring toolboxes (e.g. Sen2Agri, Sen4CAP).

To fill gaps due to cloud cover, the S2 vegetation index time series were temporally interpolated using the Image Time Series Gap Filling tool (Inglada, 2016) available in Orfeo Toolbox (Grizonnet et al., 2017). As we intend to apply this classification to a larger area, the time series were temporally resampled to a 5-day grid, starting at the first acquisition date, to overcome the multiple-day offset between adjacent satellite tracks (Inglada et al., 2015). Both linear and cubic spline interpolations were tested for the three indices. A total of 6 different S2 feature sets were thereby tested as input features for the random forest classifier (Table 4.1).

Microwave data guarantee regular temporal coverage and can provide complementary information to optical data. The complementarity of Sentinel-1 and Sentinel-2 has been shown in the context of grassland mowing detection (Chapter 3). Therefore, we tested the classification with microwave time series alone and in combination with Sentinel-2 data. Sentinel-1 γ_0 backscattering amplitudes in VV and VH polarization and the ratio VV/VH were used as input features. Ascending and descending pass acquisitions are made at different times of the day and with different look angles. Since radar signal is strongly impacted by water content, morning acquisitions are significantly affected by dew and vegetation water content. Each polarization was therefore tested in ascending (e.g. VV_{asc}) and descending (e.g. VV_{desc}) pass separately. A total of 6 different S1 feature sets were thereby tested as input for the classification (Table 4.1).

In addition, to assess the complementarity of S1 and S2 for differentiating *pastures* and *hay meadows*, the best-performing S2 feature set was tested in combination with the best-performing S1 feature as well as the respective time series minimum (*min*), maximum (*max*), mean (*mean*) and median (*median*) value and all statistics together (*stats*). A total of 6 different S1 and S2 feature combinations were thereby tested.

Table 4.1: Sentinel-1 and 2 images acquired during the study period (April 9 to July 19, 2019) and used to compute the classification features.

	Max(cloud)	Pass/Tile	n	Features
S1	-	ascending	17	VV_{asc}, VV_{desc}, VH_{asc},
		descending	18	VH_{desc}, $ratio_{asc}$, $ratio_{desc}$
S2	80%	31UFR	17	$NDVI_{linear}$, $NDVI_{spline}$, $CIre_{linear}$, $CIre_{spline}$, LAI_{linear}, LAI_{spline}

Classification mask

The classification mask was built by combining and resampling a grassland mask and a shadow mask (Figure 4.2).

The grassland mask was obtained by reclassifying the 2 m resolution land cover product of LifeWatch (Radoux et al., 2023). Two grassland classes, namely "Monospecific grassland with graminoids" and "Diversified grassland and shrubland" were taken into account. The first are herbaceous covers that are not plowed in the ongoing year, including permanent and temporary agricultural grasslands. The second class includes herbaceous covers with higher ecological value, which were defined by visual interpretation.

The shadow mask was based on a digital surface model (DSM) (Figure 4.3). The DSM of Wallonia is a product of the orthophoto acquisition campaign of 2019 (Service Public de Wallonie, SPW). Shadow projections were computed with 2m resolution based on the object heights from the DSM and a sun azimuth and elevation of 146°and 38°respectively.

The combined grassland and shadow mask was resampled to 10m to match Sentinel-2 pixels. A minimum rule was applied for the resampling to take only pure pixels into account for the classification, with 100% grassland and no shadow.

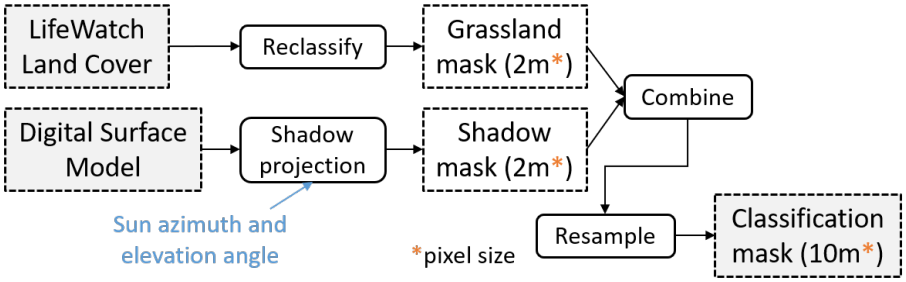

Figure 4.2: Classification mask flowchart. The mask was built using the two grassland classes of the LifeWatch land cover (Radoux et al., 2023) product and shadow projections based on a digital surface model (DSM). The mask was resampled to 10m to match the pixel size of Sentinel-2.

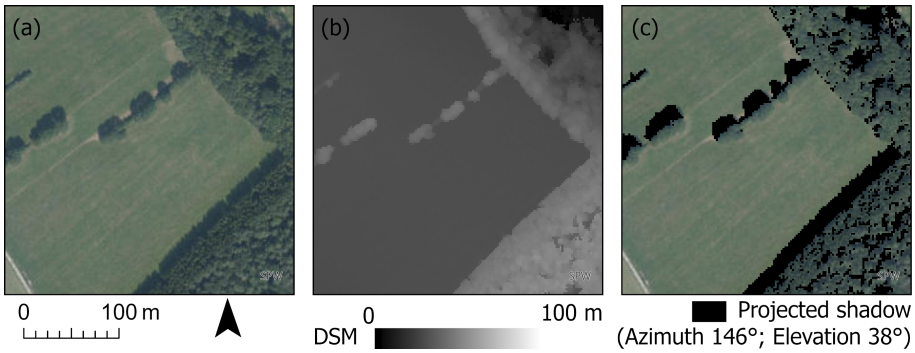

Figure 4.3: a. orthophoto (SPW) with tree shadows in a grassland, b. Digital Surface Model (DSM, credits SPW) showing the elevation of the top of objects above the ellipsoid and c. projected shadows based on the DSM with given sun angles.

This mask allows to classify only pure grassland pixels and discard pixels that are influenced by shadows or trees (Figure 4.4). The LAI time series of masked pixels (in grey on Figure 4.4) consistently differ from the valid pixels (in green on Figure 4.4). The majority of masked pixels in this example are mostly influenced by shadow, which manifests in lower LAI values throughout the season. A few masked pixels are influenced mostly by trees and shrubs and have higher LAI values compared to valid grassland pixels.

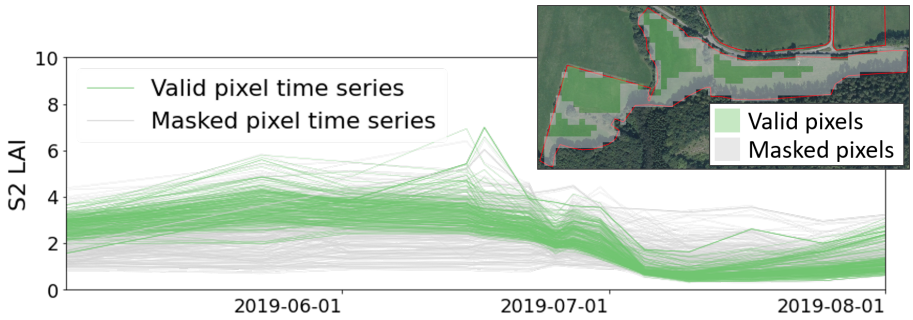

Figure 4.4: Leaf Area Index (LAI) time series per valid (green) and masked (grey) pixel for a grassland parcel (drawn in red on the orthophoto (SPW)).

Reference data

Based on the field observations, the observed parcels could be classified into two categories: *pastures*, that were exclusively grazed during the study period, and *hay meadows* on which at least one mowing event was observed. The reference parcels were redrawn manually, based on the LPIS, the grassland mask, and the Walloon orthophoto of 2019 (SPW) to obtain homogeneous reference parcels. When two or more management units could be differentiated inside one declared parcel, only the management units closest to the road was considered to be matching the field observation.

During the redrawing, 5 parcels were discarded because they contained no valid pixels due to shadow. In total, the reference dataset contained 412 parcels (194 *pastures* and 218 *hay meadows*). They were equally partitioned into a training and a validation dataset through stratified random sampling. The training dataset was used to train, calibrate and compare the classification methods through cross-validation. The validation dataset was used to validate the final product.

Cross-validation

Different classifiers were evaluated and compared through a 4-fold cross-validation scheme. The training dataset was split into 4 subsets

to keep a reasonable number of samples for the validation at each iteration. The classifiers were compared based on the mean overall accuracy (OA) and its standard deviation. The best-performing classifier was then trained using the whole training dataset. The resulting classification was then validated with the validation dataset. The user and producer accuracy (UA and PA) of both classes were also computed in addition to the overall accuracy.

Both during the cross-validation and the final validation, we applied a per-pixel wall-to-wall validation, assessing each pixel inside each redrawn homogeneous reference parcel.

4.3.2 Object based mowing detection

The pixel-based classification obtained in the first step was used in combination with the LPIS to obtain homogeneous parcels for an object-based mowing detection using the Sen4CAP toolbox v3.0 (Bontemps et al., 2020).

Classification post processing

The following steps were applied to obtain homogeneous grassland parcel polygons based on the classification and the LPIS.

1. The classification is filtered to remove isolated pixels: a pixel value is changed to the other class if there are less than 4 pixels of the same class in a 3×3 window around the pixel.

2. All parcels declared as grasslands (temporary or permanent) are extracted from the LPIS.

3. The LPIS grassland polygons are rasterized at 10 m resolution using the parcel unique feature ID's as raster values.

4. The binary grassland classification ($pastures$ vs $haymeadows$) and the rasterized LPIS are merged by adding them up.

5. The merged raster is polygonized

6. No-data polygons (i.e. covering masked areas) and Polygons with an area smaller than 1000 m^2 (10 pixels) are discarded

Mowing detection method

The mowing detection method of Sen4CAP is based on two separate algorithms detecting changes in Sentinel-2 and Sentinel-1 time series extracted per parcel (Figure 4.5). The detailed method is described in Chapter 3.

The S1 algorithm detects significant increases in VH interferometric coherence compared to the linear fit of the six previous values. It is based on a Constant False Alarm Rate (CFAR) adaptive threshold ($3.0 \times 10^{-7} \times \sigma$) that takes into account the standard deviation of the residual fitting errors (σ). The S2 algorithm detects a mowing event when the decrease in NDVI between two consecutive cloud-free acquisitions is larger than 0.12.

A confidence level is computed for each detection with lower values for S1 than for S2, to compensate for the lower precision of S1 mowing detection. For each parcel, the four most confident detections are retained. For each detection, the detection interval is given along with the confidence level and the data source (S1, S2, or both). The confidence levels of the detections range from 0 to 1 and are well correlated to the precision of the detections (De Vroey et al., 2022).

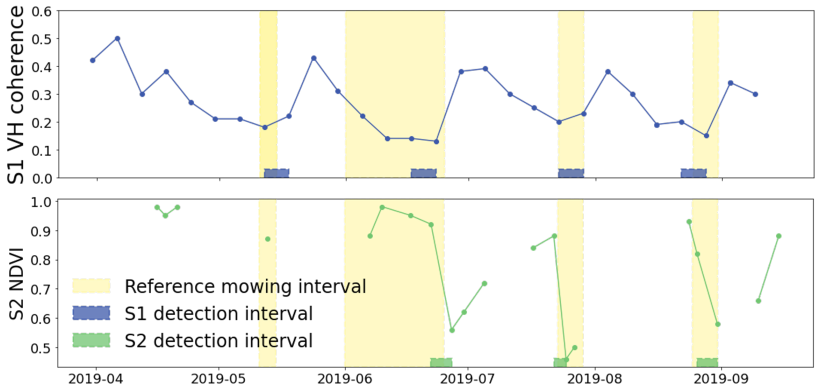

Figure 4.5: Illustration of the mowing detection algorithms of the Sen4CAP toolbox, based on Sentinel-1 VH coherence and Sentinel-2 NDVI time series extracted on a permanent grassland parcel.

Validation

The mowing detections were validated by crossing the detection intervals with reference intervals to compute the *precision* and the *detection rate*, like in previous chapters (cf. Section 2.4.3).

In addition, the accuracy of the differentiation between grasslands with an early first mowing event and a late first mowing event was validated through a confusion matrix and related quality metrics (UA, PA, and OA).

The calibration reference dataset was used to define the optimal confidence level thresholds and maximize the accuracy of the management practice classification. The validation dataset was then used to assess the result.

Here as well, to stay consistent with the previous classification validation, a per-pixel wall-to-wall validation was applied.

4.4 Results

4.4.1 Classification algorithm calibration

The results of the random forest classifier calibration with the different feature sets are shown in Tables 4.2-4.4. For each tested feature set, the mean overall accuracy (mean OA) and its standard deviation (std OA) over the 4 iterations of the cross-validation scheme are given.

Table 4.2 shows the calibration results for the Sentinel-2 feature sets. The highest mean OA is 88.4% obtained with LAI_{spline} and the lowest is 85.7% obtained with $NDVI_{linear}$. For all three indices, the cubic spline interpolation seems to result in slightly better performances than the linear interpolation. Given the standard deviation of the OA, ranging from 2.6% to 5.3%, the differences in OA between feature sets are however relatively low.

Table 4.3 shows the calibration results for the Sentinel-1 feature sets. The highest mean OA are obtained with the VV polarization, both in ascending (71.3%) and descending pass (68.5%). The lowest performances are obtained with the VV/VH ratio in both passes (60.3% and 57.7% OA). The performances are overall significantly lower than with the S2 features. Moreover, the standard deviations of the OA over the iterations are higher (4.6% to 8.7%) showing a higher sensitivity of the algorithm to the training dataset.

Table 4.2: Classification algorithm calibration results with the Sentinel-2 features. For each feature time series, the mean overall accuracy (mean OA) and its standard deviation (std OA) over the 4 iterations of the cross-validation scheme are given.

S2 features	mean OA	std OA
LAI_{spline}	88.4%	3.5%
$CIre_{spline}$	87.3%	3.0%
LAI_{linear}	87.0%	3.7%
$CIre_{linear}$	86.9%	4.8%
$NDVI_{spline}$	86.3%	2.6%
$NDVI_{linear}$	85.7%	5.3%

Table 4.3: Classification algorithm calibration results with the Sentinel-1 features. For each feature time series, the mean overall accuracy (mean OA) and its standard deviation (std OA) over the 4 iterations of the cross-validation scheme are given.

S1 features	mean OA	std OA
VV_{asc}	71.3%	6.5%
VV_{desc}	68.5%	8.7%
VH_{desc}	67.3%	7.1%
VH_{asc}	67.0%	4.7%
$ratio_{asc}$	60.3%	4.6%
$ratio_{desc}$	57.7%	5.2%

Finally, table 4.4 shows the results of the combined S2 and S1 feature sets. The best-performing S2 feature time series (LAI_{spline}) was combined with the best-performing S1 feature (VV_{asc}) to test if they improve the classification performances. The last column (delta OA) shows the change in mean OA compared to the use of S2 LAI_{spline} time series alone. The addition of the VV_{asc} time series to LAI_{spline} improves the mean OA by 0.2%. The VV_{asc} temporal statistics all result in a small decrease in performance, from -0.1% with $max(VV_{asc})$ to -1.1% with $min(VV_{asc})$. Overall, the differences in mean OA and std OA are extremely small compared to those of S2 LAI_{spline} alone.

The calibration results suggest that the three tested S2 vegetation index time series allow differentiating pastures from other grasslands with high accuracy. S1 γ_0 backscattering amplitude time series how-

Table 4.4: Classification algorithm calibration results with the combined S1 and S2 features. For each feature time series, the mean overall accuracy (mean OA) and its standard deviation (std OA) over the 4 iterations of the cross-validation scheme are given. The change in OA compared to the use of S2 LAI_{spline} (delta OA) is provided in the last column.

S2 feature	S1 features	mean OA	std OA	delta OA
	VV_{asc}	88.6%	3.0%	+0.2%
	$max(VV_{asc})$	88.3%	3.6%	-0.1%
LAI_{spline}	$mean(VV_{asc})$	88.2%	3.8%	-0.3%
(OA=88.4%)	$stats(VV_{asc})$	87.9%	3.0%	-0.5%
	$median(VV_{asc})$	87.6%	3.7%	-0.9%
	$min(VV_{asc})$	87.3%	4.1%	-1.1%

ever performed lower. Moreover, the combination of S1 VV_{asc} time series or temporal statistics to S2 features did not significantly improve the classification accuracy. For further analysis, we used the S2 LAI time series with cubic spline interpolation since it provided the highest mean OA of the S2 features.

4.4.2 Classification validation and post-processing

Based on the results of the calibration, a new classifier (Random Forest with MaxDepth 20) was trained using the whole training dataset (n=208) and applied to the Sentinel-2 LAI time series with spline interpolation. The resulting classification is shown in Figure 4.6. Visually, the classification seems relatively accurate in separating *pastures* from *haymeadows*. In some parcels, there is however a salt-and-pepper effect due to the pixel-based approach. The classification was quantitatively validated using the remaining half of the reference dataset (Table 4.5). The overall accuracy is 88%, which is very close to the OA obtained during calibration (88.4%). The user and producer accuracies are also high and well-balanced. The UA is 88% for both classes. The PA is 91% for the *pastures* and 85% for the *hay meadows*.

To assess the added value of masking out pixels containing non-grassland elements (e.g. trees or buildings) or shadows, the same classifier was applied using only the grassland parcels from the LPIS as a

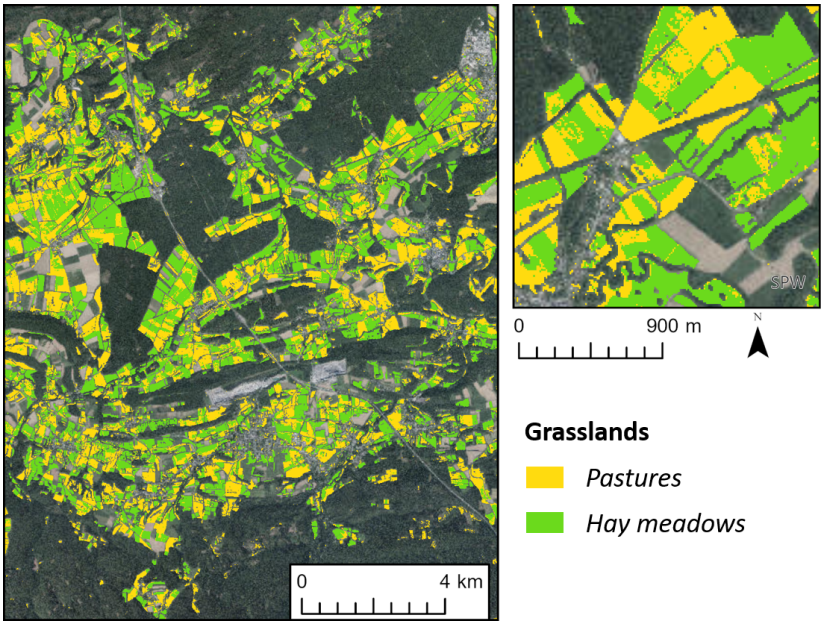

Figure 4.6: Pixel-based classification based on Sentinel-2 LAI time series differentiating *pastures* (exclusively grazed) from *hay meadows* (including mixed practices).

classification mask. The OA of this classification is 87%, which is 1% lower than when using high-resolution products (land cover and DSM) to build a strict classification mask. In the examples in figure 4.7, the classification without LC and shadow mask shows some commission errors due to trees and shadows inside parcels, while these pixels are masked out when using high-resolution products. The impact on the OA is small since only a limited number of pixels (9%) are involved.

The obtained classification was post-processed to obtain homogeneous grassland management units with a single management practice. The removal of isolated pixels (cf. section 4.3.2) allowed to significantly reduce the salt and pepper effect (Figure 4.8 (b)). The filtered raster was then crossed with the LPIS raster and polygonized. Polygons smaller than 1000 m^2 (10 pixels) were discarded to obtain clean homogeneous grassland parcels for the object based mowing detection (Figure 4.8 (c)). From the 20 796 grassland parcels declared in the LPIS in the study area, 21% contained both *pasture* and *haymeadow* manage-

Table 4.5: Validation of the pixel-based grassland classification differentiating *pastures* (exclusively grazed) from *hay meadows* (including mixed practices). Confusion matrix, user, producer, and overall accuracy (UA, PA, OA) between reference (ref) and predicted (pred) types.

ref\pred	*pastures*	*hay meadows*	**PA**	**UA**
pastures	24267	2516	**91%**	**88%**
hay meadows	3277	18843	**85%**	**88%**
			0A	**88%**

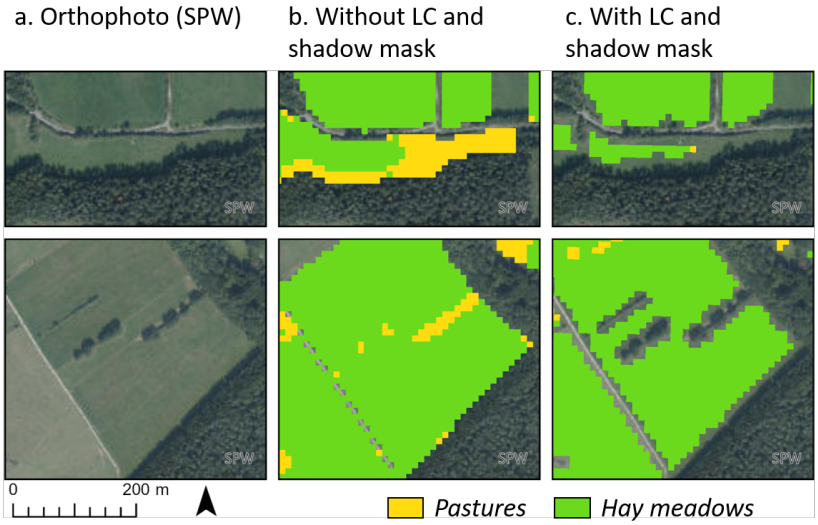

Figure 4.7: Comparison of the classification obtained with (c) and without (b) the land cover (LC) and shadow mask derived from very high-resolution data.

ment units. Furthermore, 10% of the remaining parcels contain multiple management units with the same practice (mowing or grazing). In total, 31 230 homogeneous grassland parcels were delineated through the classification in the study area.

4.4.3 Mowing detection calibration and validation

The object-based grassland mowing detection method of the Sen4CAP toolbox was applied to the polygons that were classified

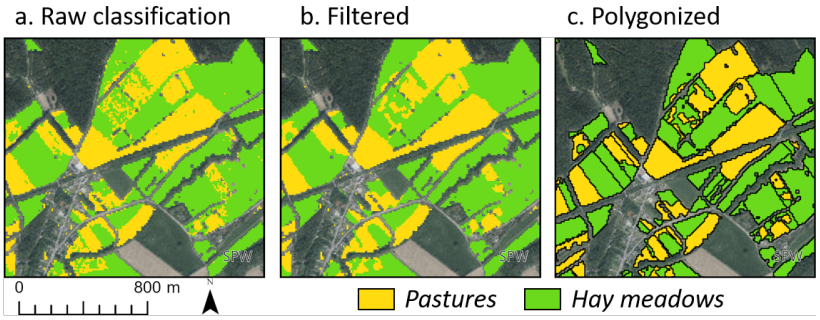

Figure 4.8: Post-processing of the raw classification result (a). Filtered to remove isolated pixels (b), polygonized and cleaned to remove small parcels (c).

as *hay meadows*. The mowing detections between April 9th and July 19th were validated with the reference mowing intervals from the calibration reference dataset. The obtained *precision* is 83% and the *detection rate* is 73%.

Based on the date of the first mowing event (before or after June 15th) two classes of mown grasslands could be differentiated: *early* and *late*. The confusion matrix and accuracy metrics computed with the calibration reference dataset are given in Table 4.6. When all detections are taken into account (minimum confidence level ($min(conf)$) = 0.0), the OA is 61%. More than half of the *late* grasslands are incorrectly classified as *early* (PA = 45%) and the UA of *early* grasslands is only 51%. This is due to false detections occurring before 15/06.

The confidence level was used to filter out early false positives. With a $min(conf)$ of 0.5 for detections before 15/06 and 0.4 for detections on or after 15/06, the PA of *late* grasslands and the UA of *early* grasslands are significantly higher (74% and 62% resp.). There are more omissions of early mowing events, reducing the PA of the *early* grasslands (57%) and the UA of the *late* grasslands (75%). However, the performances are more balanced and the overall accuracy is higher (67%). This adaptive $min(conf)$ was therefore retained for further grassland characterization.

The Sen4CAP mowing detection on *hay meadows* (with the adaptive $min(conf)$) was finally validated with the independent validation reference dataset. The estimated *precision* of the detections is 93% and

Table 4.6: Calibration of the minimum mowing detection confidence level ($min(conf)$). Confusion matrices crossing the first mowing event classes (before ($early$) or after June 15th ($late$)) based on field observations (ref) and mowing detection (pred) for different $min(conf)$ of detections. The User and producer accuracies (UA and PA) are provided as well as the overall accuracy (OA).

$min(conf)$ = **0.0**				
ref\pred	*early*	*late*	no activity	**PA**
early	6994	1262	124	**83%**
late	6389	5545	326	**45%**
UA	**51%**	**81%**	OA	61%
$min(conf)$ = **0.5 (< 15/06), 0.4 (≥ 15/06)**				
ref\pred	*early*	*late*	no activity	**PA**
early	4744	3033	603	**57%**
late	2871	9063	326	**74%**
UA	**62%**	**75%**	OA	67%

the *detection rate* is 82%. According to the validation dataset, grasslands with *late* and *early* first mowing events were differentiated with an overall accuracy of 75%.

4.4.4 Hierarchical classification of management practices

Based on the previous results we can expect that the classification and mowing detection allow to hierarchically differentiate three grassland management practices with high accuracy. First, *pastures* are differentiated from *hay meadows* through the classification. Then the *hay meadows* can be further differentiated by the first mowing date (before or after June 15th). This hierarchical classification was validated using the validation reference dataset (Table 4.7). The overall accuracy is 79%. The UA and PA of the *pasture* class (resp. 89% and 91%) are slightly improved compared to the raw pixel-based classification (Table 4.5), thanks to the post-processing. The UA and PA of the *hay meadows* with *early* (resp. 65% and 54%) and *late* (resp. 70% and 72%) first mowing are lower due to the confusion between both sub-classes.

In addition to the first mowing date, *hay meadows* can be differentiated by the number of mowing events in the growing season. *hay meadows* with an early first mowing (< 15/06) were further split

Table 4.7: Validation of the hierarchical grassland typology. Confusion matrices crossing the main management practice classes (*pastures* and *hay meadows*) and the first mowing event classes (before (early) or after June 15th (late)) based on field observations (ref) and on the classification and mowing detection (pred). The User and producer accuracies (UA and PA) are provided as well as the overall accuracy (OA).

ref\pred		*pastures*	*hay meadows*			**PA**
			early	late	no activity	
pastures		24224	492	1397	398	**91%**
hay meadows	early	1523	4846	2588	64	**54%**
	late	1587	2082	9255	0	**72%**
UA		**89%**	**65%**	**70%**	OA	**79%**

into grasslands with less than 3 mowing events (n < 3) and 3 events or more (n ≥ 3), while those with a late first mowing (≥ 15/06) were further split into grasslands with only one mowing event (n = 1) and 2 events or more (n ≥ 2). These final grassland classes are mapped in Figure 4.9.

4.5 Discussion

4.5.1 Classification and mowing detection performances

One of the main motivations behind the binary classification developed in this chapter was to be able to exclude pastures for subsequent mowing detection. In previous studies, grazed parcels were either not taken into account (Lobert et al., 2021) or shown to be a confounding factor for mowing detection (Griffiths et al., 2020; De Vroey et al., 2022; Schwieder et al., 2022). Precise information on the management practice of grasslands (i.e. mowing or grazing) is however rarely available. Using a large field data set, we showed that this information could be retrieved with high accuracy from Sentinel-2 vegetation index time series. This corroborates the hypothesis that grazed grasslands can be distinguished from mown grasslands based on their relatively constant temporal vegetation index profiles. The LAI had already been identified as a relevant variable to discriminate grazed and mown grasslands in a study using three SPOT images (Dusseux et al., 2014).

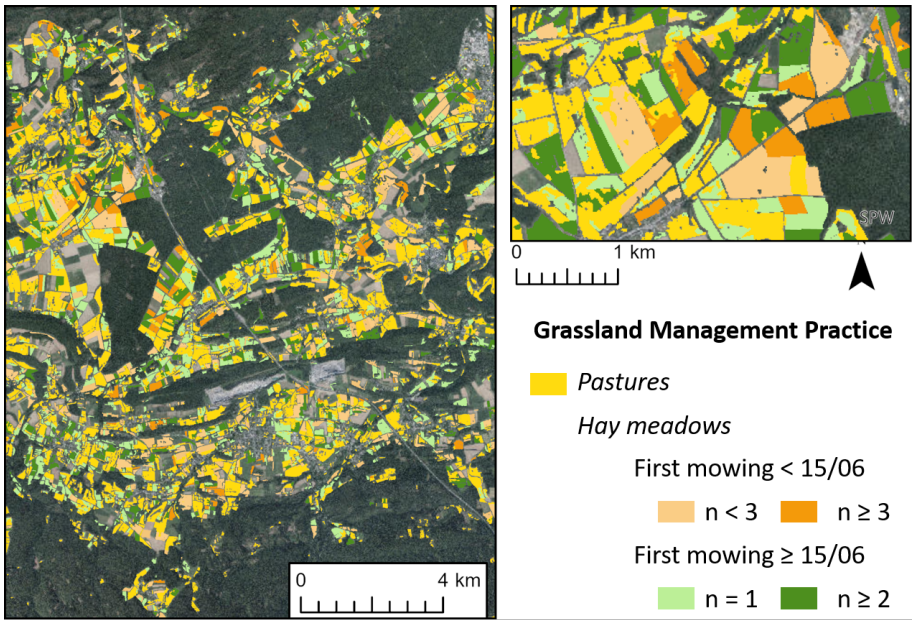

Figure 4.9: Grassland management practice classification. This hierarchical classification is based on the classification differentiating *pastures* from *hay meadows* and the mowing detection which further differentiates the second class by the date of the first mowing and by the number of mowing events (n).

In this study, the three tested vegetation indices derived from Sentinel-2 (NDVI, CIre, and LAI) performed similarly and the random forest classifiers all reached high overall accuracy. The performances obtained with Sentinel-1 backscattering time series were much lower. This can mainly be explained by the speckle effect inherent to SAR imagery that makes pixel-based analysis challenging without any spatial or temporal smoothing. The addition of Sentinel-1 backscattering temporal statistics to Sentinel-2 input features did not significantly improve the classification results. Sentinel-1 was therefore discarded for the classification step. The LAI time series with cubic spline interpolation was retained for further analysis since it performed slightly better, but the NDVI and the CIre could be used as well since the differences in performances were not statistically significant.

Another, related, aim of this pixel-based classification was to tackle the issue of grassland parcel delineation, raised in previous chapters and illustrated in Figure 4.1. In datasets such as the LPIS, parcel delineations often include several management units that are managed differently or at different times. The binary classification and the post-processing including a filtering step to remove isolated pixels, allowed to retrieve more homogeneously managed grassland patches at the management units level.

Next to the heterogeneity of practices, delineated grassland parcels can also include hedges, trees, and buildings with different spectral signatures that can hinder the classification. Thanks to the 2 m resolution land cover product that was used to build the grassland mask, the 10 m pixels with less than 100% grassland cover could be masked out. In optical remote sensing, shadows can also be a significant issue. A shadow mask, estimated through a DSM was therefore added to the grassland mask to further optimize the classification performances. Due to the shadow mask, part of the grassland area (9%) was discarded from the classification. The shadows are mostly located at the border of the parcels. Under the assumption that it is managed identically than the core area of the field, the lost grassland area could therefore be recovered by dilating the final classification and masking it again with the grassland mask alone. Overall, the availability of very high-resolution products such as the land cover map, the orthophoto, and the DSM was a great asset. Very high-resolution data and products are increasingly available and could be used to build similar grassland masks and reproduce the classification over larger areas.

The operational object-based mowing detection method of the Sen4CAP toolbox was applied to the homogeneous patches of *hay meadows* retrieved from the classification. According to the validation reference dataset, the method reached a *precision* of 93% and a *detection rate* of 82%. These detection performances are much higher than those obtained on the same grasslands without the preliminary classification, especially in terms of *precision*. The *precision* was only 44% when the pastures were taken into account, due to false mowing detections on grazed grasslands (Chapter 3). The exclusion of pastures thanks to the classification was of course a major factor in this increased performance. However, even compared to the *precision* we obtained in Chapter 3 on hay meadows alone (73%), the present results show a sig-

nificant improvement. This implies that the homogeneity of practices and the absence of trees and shadows inside the reshaped grassland parcels also contributed to the high mowing detection accuracy. In addition, the wall-to-wall pixel-based validation applied in this chapter could also explain the higher performance metrics, since the size of the parcels was not taken into account in the validation in the previous chapter. The mowing detection performances obtained here are also slightly higher than those obtained with a deep learning approach combining Sentinel-1, Sentinel-2, and Landsat-8 in a convolutional neural network with a maximum *precision* of 86% and a *detection rate* of 82% (Lobert et al., 2021).

While this method showed high performances on managed grasslands in our study area in the 2019 growing season, it should be further tested in more extended areas and other seasons. For example, the effects of drought on vegetation could significantly alter the vegetation index time series and thereby represent a challenge for classification and mowing detection. Furthermore, there are some grasslands that are neither mown, nor grazed. This study focused on managed grasslands. Unmanaged grasslands were not included in the reference dataset, as they are relatively rare in the study area. With a more complete and diverse reference dataset, this approach could be tested with more grassland management classes.

4.5.2 Grassland typology and perspectives

Previously, a few studies have considered the classification of grassland management practices and intensities through remote sensing, showing promising results, but often lacking sufficient representative ground truth data for validation. Using a supervised classification algorithm on RapidEye imagery and a rule-based method to estimate the first mowing date, Franke et al. (2012) classified four types of grassland (semi-natural, extensive, intensive and tilled) with high accuracy on a small study area in Germany. The red edge vegetation index derived from 5 RapidEye images was used by Gómez Giménez et al. (2017) to retrieve a grassland use intensity index based on the individual estimation of three factors (mowing, grazing, and fertilization intensity). They obtained promising results for the estimation of graz-

ing and mowing intensities but lacked actual ground truth data for validation.

In this study, the hierarchical categorization based on the classification and the mowing detection allowed to differentiate five types of grassland based on the main management practice (grazing or mowing), the date of the first mowing event, and the mowing frequency. Thanks to the large and regionally representative field dataset, we showed that three classes (pastures, meadows with an early first mowing event, and a late first mowing event) could be differentiated with 79% overall accuracy. The mowing frequency estimation could not be validated since the field campaign was only carried out between April 9th and July 19th, while mowing events occur until the end of October. However, given the high detection accuracy obtained during the study period, we make the hypothesis that the detections remain relatively accurate throughout the season.

While *hay meadows* could be further differentiated through the mowing detection, *pastures* were not further categorized. In a recent study with a similar hierarchical categorization approach, pastures and mown grasslands were differentiated based on biomass productivity and both classes were then subdivided into 3 intensity levels based on the exploitation (i.e. harvest) frequency (Stumpf et al., 2020). Both the biomass productivity and the exploitation frequency were retrieved through the detection of significant drops in Landsat NDVI time series, considering the cumulative change and the count of drops respectively. While this approach showed consistent results with regional statistics and georeferenced land use data, the land use intensity levels of both classes could not be validated due to a lack of ground truth data. Moreover, the timing of the first exploitation activity should be considered in addition to the exploitation frequency as it is a major factor of grassland use intensity and has an influence on their ecological value (Humbert et al., 2012; Johansen et al., 2019).

We showed that the retrieval of homogeneously managed grassland patches and the identification of pastures greatly improved the precision of mowing detection and allowed to classify five grassland types with high accuracy. These management units could further serve as a baseline to retrieve other grassland characteristics and study their relationships with biodiversity and ecology.

This hierarchical approach, combining different spatial units and image analysis methods, could be applied in broader contexts. More grassland classes could be included in the initial pixel-based classification (e.g. unmanaged grasslands, different types of vegetation covers ...) and then further differentiated through object-based time series analyses. It should be further developed in different conditions to be able to extend it over larger areas and transfer it to other seasons to classify grasslands at the landscape level (Johansen et al., 2019; Shahan et al., 2017) and study inter-annual variations (Allan et al., 2014) to contribute to ecological habitat monitoring.

4.6 Conclusions

Several studies have shown the great potential of remote sensing for grassland monitoring. In particular, the most recent developments in automated mowing detection methods allow estimating mowing dates and frequencies with high accuracy. In this chapter, we built on previous achievements to produce a thematically improved grassland classification, differentiating five management classes. First, a pixel-based classification using LAI time series differentiated *pastures* from *hay meadows* with an overall accuracy of 88%. An object-based mowing detection method using Sentinel-1 coherence and Sentinel-2 NDVI was then applied to further differentiate *hay meadows* by timing and frequency of mowing events. The pixel-based approach and the strict grassland mask built for the classification allowed to retrieve homogeneous grassland management units. Moreover, the preliminary identification of grazed grasslands reduced the number of false positives due to confusion between grazing and mowing activities during the mowing detection. The hierarchical classification method differentiated pastures, and meadows with an early first mowing event and a late first mowing event with an overall accuracy of 79%. The retrieved management practices could be combined with other factors and environmental context for further grassland characterization to contribute to ecological habitat monitoring.

Chapter 5

Retrieving grassland nitrogen nutrition status and forage yield from Sentinel-2

This chapter will be submitted as:

De Vroey, M., Radoux, J., Farinelle, A., & Defourny, P. Retrieving grassland nitrogen nutrition status and forage yield from Sentinel-2. In preparation.

Abstract

Comprehensive information on grassland yield is crucial to evaluate the combined impacts of pedo-climatic conditions and management practices on productivity. In this chapter, the objective is to estimate and assess the retrieval performance of grasslands dry matter yield (DM, [t/ha]), nitrogen concentration (N%, [g/kg]), and canopy nitrogen content (CNC, [kg/ha]) from Sentinel-2. These S2-based biophysical variables are then combined for a large scale analysis of (i) grasslands spring vegetation status, and (ii) harvested forage yield to evaluate the impacts of management practices on grassland outputs for different agroecological regions. A stepwise linear regression approach is applied to retrieve grasslands DM, N%, and CNC from Sentinel-2 reflectances and indices. The models are calibrated using field measurements made during three growing seasons. The models were validated on an independent field dataset, achieving a normalized RMSE of 9.5%, 15.6%, and 6.7% for DM, N%, and CNC respectively. In addition, the nitrogen nutrition index (NNI) was computed from the estimated DM and N% to evaluate grasslands nitrogen nutrition status. The retrieval models were then used in combination with classified management units and mowing detections for a regional evaluation of spring vegetation status and harvested forage yield. These large-scale applications consistently showed the impacts of management practices and pedo-climatic conditions on grassland DM and N content.

5.1 Introduction

Grasslands are a key element of many agricultural systems as they provide nearly half of the feed requirements for global livestock production (O'Mara, 2012; Herrero et al., 2013). Adapting grassland management practices to sustain essential ES in a changing climate, while meeting the increasing demands of dairy and meat production constitutes a great challenge. Comprehensive information on grassland yield is crucial to evaluate the combined impacts of pedo-climatic conditions and management practices on their productivity. It can furthermore facilitate the transition towards more sustainable grassland management to create synergies and balance trade-offs with biodiversity, carbon budget, and other regulating ES.

Nitrogen (N) is a key nutrient for grassland production. N strongly influences vegetation growth, production and quality as it plays a role in carbon fixating and light harvesting complexes (Berger et al., 2020b). Intensively managed pastures and hay meadows are often fertilized with N to increase their biomass productivity. Furthermore, nitrogen content is a good indicator for the nutritional quality of forage, as it can be directly related to its protein content. For instance, comprehensive data on grassland nitrogen content could support developments toward protein self-sufficiency in livestock farming.

Based on yield and N content measurements, the nitrogen nutrition index (NNI) of grasslands can be assessed. The NNI is commonly used for fertilization management as it indicates the nitrogen saturation level of the plant. It is based on the nitrogen dilution curve, which describes the decrease in critical N concentration with increasing biomass during the vegetative growth stage. This curve has been defined empirically for several crops and for grasslands (Gastal and Lemaire, 2002).

Dry biomass yield (DM) and N content can be estimated through ground-based methods, such as visual assessment, rising plate meters, field spectroscopy, cutting, drying and laboratory analysis. These methods are, however, either very costly and time-consuming or highly subjective and cannot be applied on extended areas (Ali et al., 2016a). Therefore, biophysical variable retrieval by remote sensing offers a great potential for exhaustive large-scale forage production assessment (Ali et al., 2016a; Reinermann et al., 2020).

Several biophysical variables can be retrieved or estimated from remotely sensed spectral reflectances, either through empirical approaches (e.g. regression), deterministic approaches (e.g. radiative transfer models (RTM)), or a combination of both. In the context of agricultural monitoring by remote sensing, Weiss et al. (2020) define both DM yield and N content as secondary variables, that are not directly related to radiative transfer mechanisms. They are however related to several processes within the soil-plant-atmosphere continuum, and thus, indirectly drive the radiative transfer. Empirical approaches allow to directly derive these secondary variables from reflectances without prior assumptions, but strongly rely on training data and are therefore often limited in terms of transferability. Deterministic approaches on the other hand require a good understanding of the processes involved to make strong assumptions and build robust models. Deterministic models, such as RTM can be used to retrieve primary variables (i.e. directly related to radiative transfer mechanisms), which can in turn be related to secondary variables (Weiss et al., 2020).

The chlorophyll content is often used as a proxy for N retrieval (Baret et al., 2007; Delloye et al., 2018; Berger et al., 2020b), since a large part of N taken up by the plant is invested in chlorophyll. Canopy chlorophyll content can be linked to reflectances in the visible and red edge domain (Hank et al., 2019) and its retrieval has been broadly studied for various crops (Verrelst et al., 2013; Delloye et al., 2018). The N-chlorophyll relation however needs to be considered with caution as a larger part of N is invested in proteins, and decreases in chlorophyll are not necessarily caused by N deficiency (Berger et al., 2020b). For DM yield estimation, the LAI can be used as a primary variable (Punalekar et al., 2018; Kayad et al., 2022). In optical multi- and hyperspectral remote sensing, the LAI is mostly linked to reflectances in the red edge and near-infrared domain (Delloye et al., 2018; Hank et al., 2019).

Studies on crop nitrogen monitoring retrieve either N concentration (often as a percentage of dry biomass), or N content per leaf- or land area. Reflectance should be more sensitive to the amount of absorbers per unit land area, than to their concentration (Berger et al., 2020b). Nevertheless, both N concentration and N content per area can be used to compute the NNI (Baret et al., 2007).

While crop yield and nitrogen monitoring are broadly studied (Berger et al., 2020b), most studies on grassland biophysical variable

retrieval by remote sensing focused on green biomass (Sibanda et al., 2015; Zhang et al., 2015; Ali et al., 2016b; Quan et al., 2017; Punalekar et al., 2018; Hardy et al., 2021). Grassland above-ground biomass has been retrieved through empirical approaches such as linear regressions, random forest, or partial least square regression, mostly from NIR and red edge reflectances (Ullah et al., 2012; Sibanda et al., 2015; Dusseux et al., 2015; Hardy et al., 2021). Recently, more studies have used deterministic approaches, such as RTM to retrieve LAI, which is, in turn, used to estimate grassland biomass (Punalekar et al., 2018; Cisneros et al., 2020; Schwieder et al., 2022).

Only a few studies developed grassland nitrogen and chlorophyll content retrieval methods, mostly through field spectroscopy (Boegh et al., 2013; Adjorlolo et al., 2014; Cisneros et al., 2020; Pullanagari et al., 2021). Overall, studies on grassland biophysical variable retrieval show promising results, but methods need to be tested either over larger areas or longer periods.

In this chapter, the objective is to estimate — and evaluate the retrieval performance of — grasslands dry matter yield (DM, t/ha), N concentration (N%, g/kg), canopy N content (CNC, kg/ha), and NNI from Sentinel-2. More specifically, DM, N%, and CNC are directly retrieved from S2 spectral bands and derived indices, through stepwise multiple linear regression. Through this empirical approach, we aim at identifying the most explanatory S2 derived variables for grassland biomass yield and nitrogen content. Furthermore, we compare the direct and indirect retrieval of grassland N, respectively using N% on one hand, and CNC on the other.

Finally, these EO-based biophysical variables are combined with the grassland management classification of chapter 4 for a large-scale and comprehensive analysis of (i) spring vegetation status, and (ii) harvested forage yield to document the impacts of grassland management practices in different agroecological regions and link inputs and outputs of grassland use intensity.

5.2 Data and method

Three object-based stepwise multiple linear regression models are built to retrieve grasslands DM, N%, and CNC from S2 reflectances. These

empirical models are calibrated on field measurements from three consecutive growing seasons (2016-2018) and validated on independent measurements from a fourth growing season (2020) to test their temporal transferability.

The models are then used to retrieve the spring vegetation status and harvested forage yield of the spatially homogeneous grassland management units delineated in the previous chapter over the whole study area (672.9 km^2) for the 2019 growing season.

5.2.1 Satellite data

The biophysical variable retrieval methods developed in this study are based on S2 data. All S2 A and B top-of-atmosphere reflectance images covering the study area and period, covering five seasons (2016-2020), were downloaded and converted to surface reflectances using Sen2Cor v2.10 for atmospheric correction and the Fmask method (Zhu and Woodcock, 2012) for cloud and cloud shadow masking.

5.2.2 Field measurements

A large field survey was carried out across Wallonia, in the framework of operational in situ grassland monitoring by "Fourrages Mieux" (Farinelle, A., 2020). The survey aimed to evaluate the forage quality of permanent grasslands with AECM subsidized by the CAP. The monitored grassland parcels (n=59) include permanent grasslands with two types of AECM — high nature value grasslands (MC4, n=27) and natural grasslands (MB2, n=14) — as well as permanent grasslands without AECM (n=18) for the sake of comparison (Figure 5.1).

The AECM aim at preserving biodiversity and protecting soil and water through more extensive management. Fertilization is prohibited in MC4 and limited to one application of organic manure between June 16^{th} and August 15^{th} in MB2. The management requirements further include (i) the obligation to mow (or graze) between June 16^{th} and October 31^{st}, (ii) the preservation of 5 to 10% of untouched refuge zone, and (iii) restricted (MC4) or prohibited (MB2) use of phytosanitary products.

During three growing seasons (2016-2018), vegetetion samples were cut and measured before each mowing event on all monitored

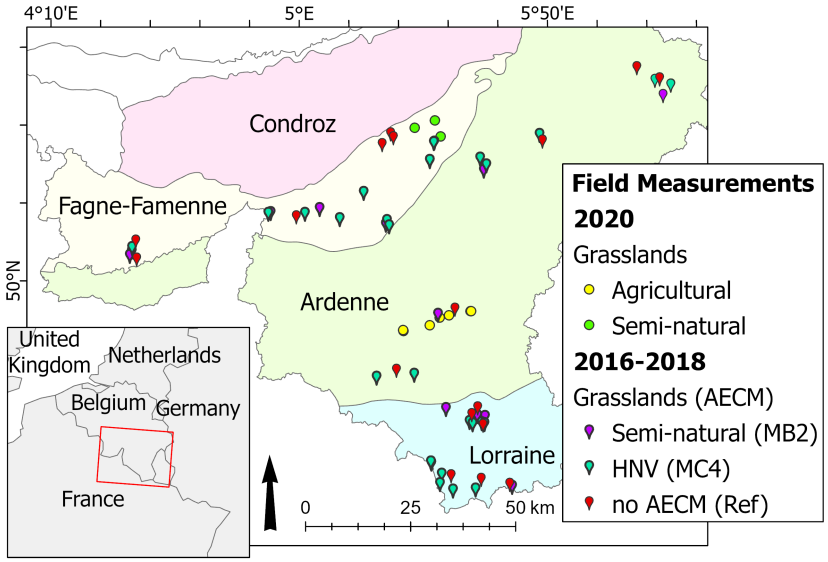

Figure 5.1: Parcels location for the 2016-2018 and the 2020 field campaign, distributed across three agroecological regions of Wallonia (Belgium).

grasslands. The same set of 59 parcels were monitored during the whole study period. The parcels were mown 1 to 5 times per season, resulting in a total of 353 measurements. For each measurement in each parcel, four samples were collected. In grasslands with AECM, the location of the samples was chosen by expert on the field to be as representative as possible of each parcel. In grasslands with no AECM, the sample points were chosen randomly on a homogeneous area of each parcel.

For each sample, an area between 6 and 12 m^2 was cut with a mower. The area was precisely measured for each sample. The cut grass was immediately weighted and a sub-sample (150-300 g) was collected, kept cool, and brought to the lab for further measurements. Each sub-sample was dried at a temperature of 55°C and weighed separately to measure the dry matter content. The parcel DM was estimated by averaging the measurements of the four sub-samples.

The N% was then measured through NIR spectrometry on a composite of the four sub-samples taken for DM measurement. The CNC

was then computed as N% × DM.

Thirteen additional DM and N% measurements were made in 2020, specifically for this study, to test the transferability of the models to another growing season. On May 11^{th}, 7 permanent grassland parcels without management restrictions were sampled and on June 25^{th}, 6 semi-natural grasslands were sampled (Figure 5.1). In each parcel, a homogeneous area of at least 20×20 m was identified and the GPS coordinates of the center of the area were recorded. Four samples were collected inside the homogeneous area, with a minimum of 4 m between each sample. The samples were collected manually, by cutting 50×50 cm squares. Similarly to the previous campaign, the collected samples were dried and weighed. The DM yield of each parcel was estimated by averaging the DM measured for each sample, the N% was measured through NIR spectrometry on a composite of the four samples, and the CNC was computed from the N% and DM.

5.2.3 Dataset preparation

Based on the field measurements an object-based reference dataset was built to calibrate and validate the regression models for the retrieval of DM, N%, and CNC.

The field measurements needed to be matched spatially and temporally to S2 spectral measurements. Each field measurement was assigned to a grassland management unit including the sample points. S2 reflectances and derived indices were averaged per management unit. The management units were delineated by combining the grassland classification and the LPIS (Chapter 4), and used as spatial units for the biophysical variable retrieval. We made the hypothesis that the delineation of these management units remains mostly unchanged through the study period (2016-2020).

The samples were not specifically collected on cloud-free S2 acquisition dates. Only 26 of the 353 measurements made in 2016-2018 could be associated with a cloud-free S2 acquisition on the same day. Therefore, the closest cloud-free S2 acquisition with a maximum interval of 9 days before the field measurement was considered useable. S2 images acquired after the field measurements could not be considered, since all samples were collected right before a mowing event. With a

maximum interval of 9 days, 154 reference points could be considered. This dataset was used to calibrate the regression models.

For the field campaign of 2020, S2 acquisitions following the field measurement could be considered as well. Cloud-free S2 acquisitions were available within 4 days around each measurement. This dataset was used to validate the regression models.

5.2.4 Stepwise linear regressions

A stepwise linear regression approach, using multiple S2 reflectances and indices, was applied to retrieve grasslands DM, N%, and CNC. Linear regression models can be calibrated with a limited number of training data, with a lower risk of overfitting, compared to more complex regression models.

Initially, all 10 m and 20 m resolution S2 bands (visible (BGR): B2-B4, red edge: B5-B7, near-infrared (NIR): B8, narrow NIR: B8A and short wave infrared (SWIR): B11-B12, Figure 1.4) and four spectral indices were considered. First, the NDVI has often been linked to green above-ground biomass (Zhang et al., 2015; Dusseux et al., 2015). The three other indices are related to the red edge, which is both linked to biomass (Sibanda et al., 2015; Schwieder et al., 2020; Hardy et al., 2021) and vegetation chlorophyll content. The CIre, the normalized difference red edge index (NDRE) and the S2 red edge position index (REP) are computed with equations 4.2, 5.1 and 5.2, respectively.

$$NDRE = \frac{B8 - B5}{B8 + B5} \tag{5.1}$$

$$REP = 705 + 35 \times \left(\frac{B4 + B7}{2 - B7} \right) / (B6 - B5) \tag{5.2}$$

A stepwise regression with a backward elimination approach was applied to select the most explanatory S2 variables for each biophysical variable. Initially, all S2-derived variables are included. At each step, a least squares linear model is fitted to the dataset and a p-value is computed for each input variable, indicating if the respective slope coefficient is significantly different from 0. The variable with the highest p-value (i.e. the least explanatory) is discarded, and a new model is fitted with the remaining variables. This process is repeated until all p-values are below a given threshold.

Three levels of variable selection were tested for each biophysical variable. The first model was calibrated without any variable selection, using all the above-mentioned input variables, the second using only variables with a p-value below 0.05, and the third using variables with a p-value below 0.01. The performances of the different variable sets were evaluated and compared through a leave-one-out cross-validation approach, using the reference dataset of the 2016-2018 growing seasons. The RMSE was computed between measured and predicted biophysical variable values. In order to be able to compare the results for DM, N%, and CNC retrieval, the RMSE was normalized by the respective biophysical variable range.

5.2.5 Independent validation

The variable selection resulting in the lowest RMSE was used to fit new models to the whole calibration dataset. These models were then used to retrieve the DM, N%, and CNC of the parcels of the 2020 field campaign to perform an independent validation. To compare the direct and indirect retrieval of N% and CNC, each N variable was also computed, based on the retrieved DM and the other N variable. The indirect N% was computed by dividing the retrieved DM by the retrieved CNC, and inversely, the indirect CNC by multiplying the retrieved N% by the retrieved DM.

In addition, the NNI was computed, from the retrieved DM and N% on one hand, and CNC on the other, to evaluate the potential of the EO-based biophysical variables to measure the nitrogen nutrition status of grasslands. The dilution curve defined for grasslands (Gastal and Lemaire, 2002) is given in equation 5.3.

$$cN\% = 4.8 \times DM^{-0.32} \tag{5.3}$$

where cN% is the critical nitrogen concentration, i.e. the nitrogen absorption capacity of the vegetation to reach maximum potential yield. The NNI is computed as the ratio between the actual N% and cN%.

5.2.6 Large-scale applications

The EO-based biophysical variables are then used to perform a comprehensive regional analysis of grasslands spring vegetation status, and the harvested forage yield. To assess the impacts of management practices, we use the classified management units and mowing detections obtained in Chapter 4 for the 2019 growing season, over an area covering three agroecological regions of Wallonia — Condroz, Fagne-Famenne, and Ardenne (Figure 2.1). Three types of management practice are considered, namely pastures, hay meadows without AECM (first mown before June 16^{th}, further referred to as "no AECM"), and hay meadows with AECM (first mown after June 16^{th}, further referred to as "AECM").

Spring vegetation status

First, the DM and N% are estimated on a single date for all management units to study the variability of the spring vegetation status cross grassland management practices. Therefore, the average DM and N% of pastures, "no AECM" hay meadows, and AECM hay meadows per region are compared.

Second, the estimated DM and N% are used to compute the NNI and map the nitrogen nutrition status of grasslands across the whole study area.

Harvested forage yield

The harvested forage yield of hay meadows is then analyzed in the three agroecological regions. Based on the mowing detections, the DM and N% are estimated before the first and second mowing events of each hay meadow management unit in the study area.

The mowing detection method (described in Chapter 3), based on S1 and S2 time series, identifies mowing events in time intervals ($[t_{start} - t_{end}]$) of maximum 18 days, in which t_{start} should correspond to the last satellite acquisition on tall grass. For each mowing event, DM and N% were estimated at t_{start}, when t_{start} corresponded to a S2 acquisition, or at the latest S2 acquisition date before t_{start}, if it was at most 18 days before t_{end}.

The average harvested forage yield (DM and N%) of hay meadows is then computed per region. The focus is set on AECM since the regulation requirements guarantee a relatively uniform timing of mowing events.

Finally, to assess the consistency of the S2-based forage yield analysis, we compare the overall trends in DM and N% assessed by field measurements on one hand, and by remote sensing on the other.

5.3 Results

5.3.1 Field measurements

The selected field measurements — including two types of grasslands and three growing seasons — are shown in Figure 5.2. On the first mowing event, grasslands with no AECM are characterized with higher N% than grasslands with AECM (Figure 5.2 (a)). At 2nd, 3d and 4th mowing events, there is no noticeable difference between AECM and no AECM.

While the measurements show no correlation between DM and N%, the CNC is strongly related to the DM, with an R^2 of 0.81 for grasslands with AECM and 0.92 for grasslands with no AECM (Figure 5.2 (b)). This is expected since the CNC is computed based on the N% and the DM.

5.3.2 Variable selection and calibration

The performances of the stepwise linear regressions, estimated through a leave-one-out cross-validation, are shown in Figures 5.3, 5.4, and 5.5.

For the DM retrieval, the best result is obtained when a p-value < 0.05 is applied for the variable selection, with an RMSE as low as 1.00 t/ha, corresponding to 15.8% of the range of measured DM values (Figure 5.3 (b)). The explanatory variables are B5, B6, B8A, B11, and the NDRE. No additional variables are discarded with a threshold p-value of 0.01. The RMSE is slightly higher when all variables are used for the regression (1.03 t/ha, 16.2%, Figure 5.3 (a)).

For the N% retrieval, a threshold p-value of 0.05 also provided the best results, with an RMSE as low as 3.32 g/kg, and a normalized RMSE of 12.1% (Figure 5.4 (b)). The explanatory variables are B4, B5, B7,

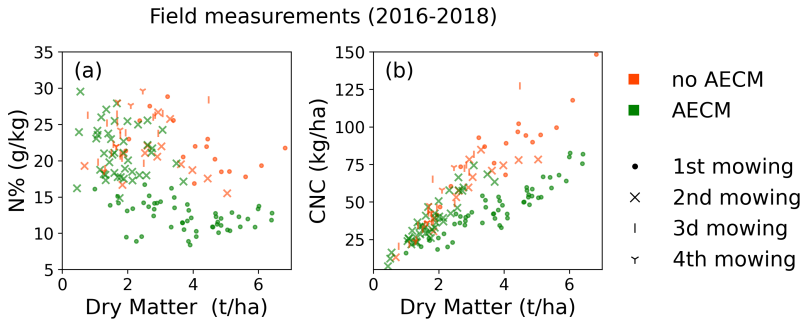

Figure 5.2: Comparison of field measurements of dry matter versus (a) N concentration (N%, g/kg), and (b) canopy N content (CNC, kg/ha) from the 2016-2018 field campaigns. The legend provides information about the AECM and before which mowing event the measurements were made.

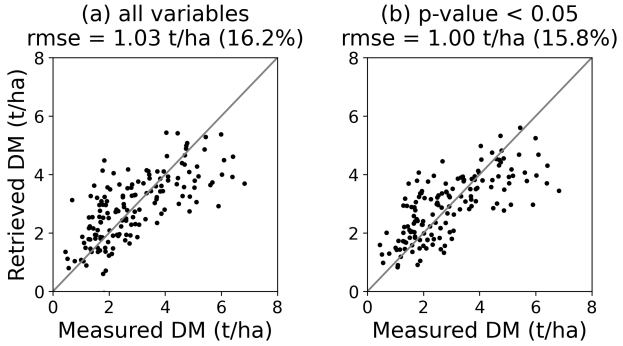

Figure 5.3: Calibration results (n=154) for DM retrieval through stepwise multiple linear regression. Including (a) all variables and (b) variables with p-values < 0.05 (B5, B6, B8A, B11 and NDRE)

B8A, B12, the NDVI, and the REP. The threshold p-value of 0.01 discarded two variables (B4 and REP), but did not improve the RMSE (3.34 g/kg, 12.2%, Figure 5.4 (c)). With all input variables, the RMSE is the highest (3.45 g/kg, 12.6%).

Finally, the CNC was also retrieved most accurately with a p-value < 0.05, resulting in an RMSE of 17.2 kg/ha, and a normalized RMSE of 12.2% (Figure 5.5 (b)), but with twice more explanatory variables (B2, B3, B4, B6, B7, B8, B11, B12, NDVI, and REP). The RMSE is slightly

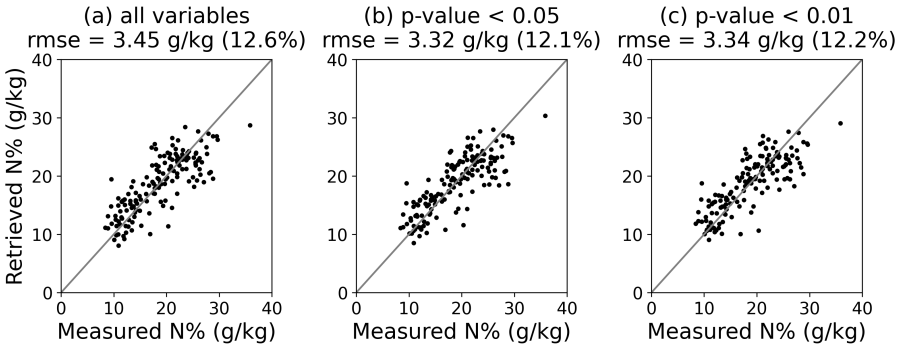

Figure 5.4: Calibration results (n=154) for N% retrieval through stepwise multiple linear regression. Including (a) all variables, (b) variables with p-values < 0.05 (B4, B5, B7, B8A, B12, NDVI, and REP) and (c) variables with p-values < 0.01 (B5, B7, B8A, B12, and NDVI)

higher with p-values < 0.01 (17.4 kg/ha, 12.4%, Figure 5.5 (c)) as well as with all variables (17.5 kg/ha, 12.4%, Figure 5.5 (a)).

For the three biophysical variables a threshold p-value of 0.05 provided the best results and was retained for the final linear regression models, which were calibrated using the whole field dataset of 2016-2018.

5.3.3 Independent validation

The calibrated regression models were successfully validated using the field dataset of 2020 (Figure 5.6 (a), (b), and (c)). The DM was retrieved with high accuracy on the 13 sampled grasslands, resulting in a smaller RMSE than in the calibration (0.61 t/ha, 9.5%). The N% was retrieved with a larger RMSE of 4.27 g/kg (15.6%). The CNC was retrieved with the highest accuracy, with an RMSE of 9.5 kg/ha, corresponding to a normalized RMSE of 6.7%. Overall, the models show good inter-annual transferability.

To test the indirect retrieval of N% and CNC, both were also computed based on the retrieved DM and the other expression of N (Figure 5.6 (d) and (e)). Based on the retrieved DM and CNC, the N% was estimated with an RMSE of 4.92 kg/ha (28.7%). Inversely, based on the retrieved DM and N%, the CNC was estimated with an RMSE of 12.3

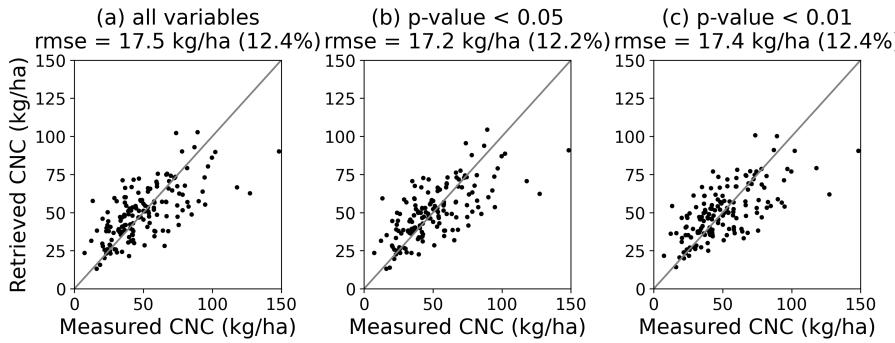

Figure 5.5: Calibration results (n=154) for CNC retrieval through stepwise multiple linear regression. Including (a) all variables, (b) variables with p-values < 0.05 (B2, B3, B4, B6, B7, B8, B11, B12, NDVI, and REP) and (c) variables with p-values < 0.01 (B4, B6, B7, B8, B11, B12, NDVI, and REP)

kg/ha (18.0%). Both for N% and CNC, the indirect and direct estimations are very similar, but the indirect estimations result in larger errors.

Based on the retrieved DM and N%, the NNI was computed and validated as well (Figure 5.6 (f)). The RMSE is 0.10 and the normalized RMSE is 14.3%.

5.3.4 Large-scale applications

The DM and N% retrieval models were used to perform a spatially comprehensive analysis of (i) grassland vegetation status in the spring and (ii) harvested forage yield. Based on the spatially homogeneous management units and mowing detections of chapter 4, the impacts of management practices could be assessed in three agroecological regions. For these regional analyzes, the N status was estimated in terms of N concentration only — and not in terms of CNC, which is closely related to the DM (Figure 5.2) — since the results of the model calibration and validation showed good performances for the direct retrieval of N%.

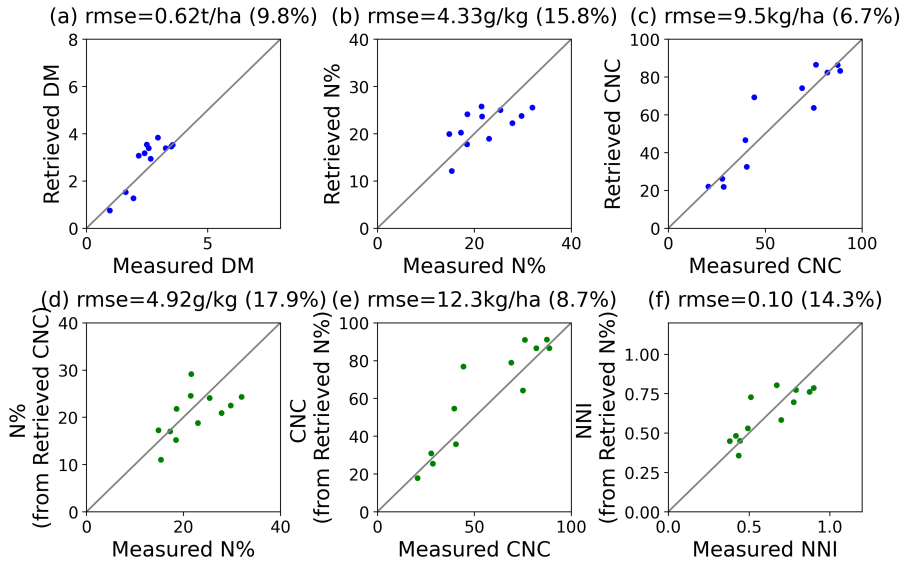

Figure 5.6: Validation of the retrieved (a) DM, (b) N% and (c) CNC, and of (d) N%, computed from the retrieved DM and CNC, (e) of CNC, computed from the retrieved DM and N%, and (f) NNI, computed from the retrieved DM and N%. The validation is based on the measurements of the 2020 field campaign.

Spring vegetation status

The DM and N% were estimated on the 1^{st} of May 2019 for all pastures and for all hay meadows that had not yet been mown at that time. The regional averages per management practice are given in Tables 5.1 and 5.2.

First, the DM in the spring is clearly driven by the pedo-climatic conditions of each region, Conroz being the most favorable, and Ardenne the least. The DM also consistently increases, as expected, according to the three grassland management categories, for Ardenne (3.01 to 3.40 t/ha), Fagne-Famenne (3.36 to 3.66 t/ha), and Condroz (3.83 to 4.17 t/ha). In each region, the DM is the highest in "no AECM" hay meadows, followed by "AECM" hay meadows and then pastures. The "no AECM" hay meadows also show the highest N% in all regions (22.0 to 22.2 g/kg). In Condroz, the N% of "AECM" hay meadows (22.1 g/kg) is however as high as that of "no AECM" hay meadows. Pastures

Table 5.1: Average (± 95% confidence interval) retrieved DM on May 1^{st} 2019 for different management practices (pastures and hay meadows with an early (no AECM) and a late first mowing event ("AECM")), per region.

Dry matter yield (t/ha) on 01/05/2019			
Grasslands: **Region:**	Pastures	Hay Meadows (AECM)	Hay Meadows (no AECM)
Condroz	3.83 ±*0.027*	3.99 ±*0.032*	4.17 ±*0.044*
F.-Famenne	3.36 ±*0.014*	3.44 ±*0.014*	3.66 ±*0.019*
Ardenne	3.01 ±*0.015*	3.17 ±*0.016*	3.40 ±*0.026*

Table 5.2: Average (± 95% confidence interval) retrieved N% on May 1^{st} 2019 for different management practices (pastures and hay meadows with an early (no AECM) and a late first mowing event ("AECM")), per region.

N% (g/kg) on 01/05/2019			
Grasslands: **Region:**	Pastures	Hay meadows (AECM)	Hay meadows (no AECM)
Condroz	21.2 ±*0.20*	22.1 ±*0.24*	22.1 ±*0.36*
F.-Famenne	20.9 ±*0.07*	21.5 ±*0.07*	22.0 ±*0.11*
Ardenne	20.9 ±*0.12*	21.4 ±*0.14*	22.2 ±*0.21*

show the overall lowest N% (20.9 to 21.2 g/kg). The variation in N% between regions is smaller. The most significant difference is observed between "AECM" hay meadows from the Condroz (22.1 t/ha) and the two other regions (21.5 and 21.4 g/kg).

The NNI was then computed from the DM and N% retrieved on May 1^{st}, to map the nitrogen nutrition status of grasslands (Figure 5.7). Some grasslands (14%) could not be considered, mostly because of a lack of S2 data due to cloud cover (13%), or because they were already mown on the 1^{st} of May (1%). It is striking to observe that the boundaries of the agroecological regions could be derived visually from the grassland NNI distributions. Condroz is characterized by more high NNI values, with 52% of the considered parcels above 0.7, compared to 29% in Fagne-Famenne and 25% in Ardenne. In each region, there are parcels with low NNI under 0.5, namely 7% in Condroz and Ardenne and 2% in Fagne-Famenne. Very few parcels have NNI above

1, which would indicate potential overfertilization (15 in Conroz, 7 in Fagne-Famenne, and 5 in Ardenne).

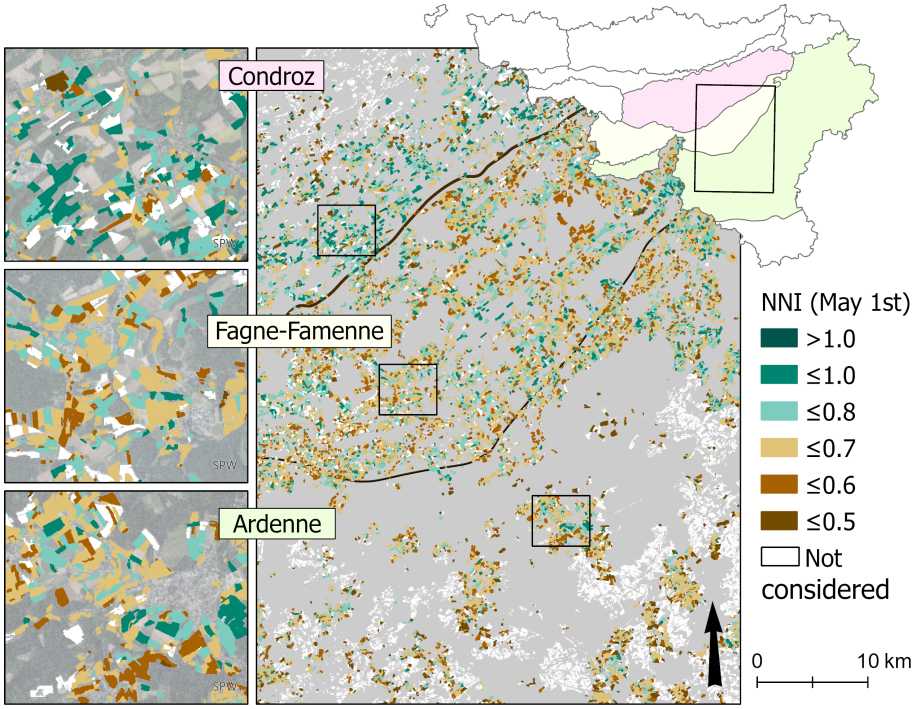

Figure 5.7: Nitrogen nutrition index computed from the retrieved DM and N% on the 1^{st} of May 2019 across three agroecological regions of Wallonia. Some grasslands could not be considered, mostly because of a lack of S2 data due to cloud cover, or because they were already mown on the 1^{st} of May. Orthophoto background for the zooms (Source: SPW).

Harvested forage yield

The average first mowing yields (DM and N%) per region for AECM hay meadows are given in Table 5.3. The average first-mowing DM is significantly higher in Ardenne (4.17 t/ha) than in Condroz and Fagne-Famenne (3.73 and 3.71 t/ha respectively).

Table 5.3: Regional means (± 95% confidence interval) of retrieved DM and N% on the first mowing event in hay meadows with AECM.

FIRST MOWING YIELD (>16/06, AECM)			
Region	n	DM (t/ha)	N% (g/kg)
Condroz	1207	3.73 ±0.05	18.6 ±0.29
F.-Famenne	3428	3.71 ±0.03	16.9 ±0.15
Ardenne	3463	4.17 ±0.03	18.0 ±0.15

In terms of N%, there are significant differences between each region, with the lowest N% in Fagne-Famenne (16.9 g/kg), followed by Ardenne (18.0 g/kg), and then Condroz (18.6 g/kg).

Finally, to assess the impacts of management practices, the first and second mowing forage yields are averaged per type of hay meadow (AECM and "no AECM"). The trend in DM and N% is assessed, from the field measurements on one hand, and from the spatially exhaustive S2-based retrieval on the other (Table 5.4).

Table 5.4: Comparison of average DM (t/ha) and N% (g/kg) (± 95% confidence interval) on mowing 1 and 2 in grassland parcels with AECM and without AECM ("no AECM"), assessed by field measurements and by remote sensing.

	MEASURED (2016-2018)		RETRIEVED (2019)	
	DM (t/ha)	N% (g/kg)	DM (t/ha)	N% (g/kg)
Mowing 1	n=50(no AECM)+111(AECM)		n= 758 + 8098	
"no AECM"	3.97 ±0.35	20.2 ±1.28	2.97 ±0.06	20.8 ±0.39
AECM	3.60 ±0.25	13.2 ±0.43	3.91 ±0.02	17.6 ±0.10
Difference	+0.37	+7.01	-0.94	+3.11
Mowing 2	n= 45 + 81		n= 3249 + 2159	
"no AECM"	2.65 ±0.29	21.2 ±1.15	2.92 ±0.03	21.2 ±0.15
AECM	1.96 ±0.15	21.5 ±0.79	1.86 ±0.03	22.8 ±0.16
Difference	+0.69	-0.39	+1.06	-1.56

Based on the field measurements, the first mowing yield of "no AECM" grasslands is characterized by a significantly higher N% (+7.01 g/kg) than with AECM. The decrease in N concentration with vegetation growth can explain the lower N% in grasslands that are mown at later development stages (AECM). This trend also shows in the S2-based N%, with a smaller, but statistically more certain difference (+3.11 g/kg), given the size of the sample (n=8856). The average

first mowing N% of grasslands with AECM is higher in the 2019 retrievals (17.6 g/kg), compared to the 2016-2018 field measurements (13.2 g/kg).

The average DM measured on the first mowing in "no AECM" grasslands is slightly higher than in grasslands with AECM (+0.37 t/ha). This 10% difference is however relatively small compared to the intra-class variability and is statistically less certain (confidence intervals ±0.35 and ±0.25). Based on the 2019 retrievals, DM yield is significantly lower on "no AECM" grasslands (-0.94 t/ha, confidence intervals ±0.06 and ±0.02). It is, however, worth mentioning a possible bias due to consistent cloud cover in the spring of 2019, preventing the retrieval of the first mowing yield for 90% of the "no AECM" grasslands. Moreover, on most (82%) of the 758 usable "no AECM" parcels, the yield was estimated before May 1^{st}, while the first mowing field measurements on "no AECM" parcels were all done after May 4^{th}. This can explain the lower average S2-based DM on the "no AECM" grasslands (2.97 t/ha), compared to the field measurements (3.97 t/ha).

For the second mowing event, the field measurements show significantly higher DM on "no AECM" grasslands (+0.69 t/ha). The S2-based variables show a larger difference (+1.06 t/ha), as the average DM of "no AECM" grasslands is higher (2.92 t/ha) and the average DM of grasslands with AECM is lower (1.86 t/ha), compared to the field measurements (2.65 and 1.96 t/ha respectively). The measured second mowing N% show no statistically significant difference between "no AECM" and "AECM" grasslands (-0.39 g/kg, confidence intervals ±1.15 and ±0.79). Based on the retrieval, the second mowing N% of "no AECM" grasslands is however significantly lower than that of "AECM" grasslands (-1.56 g/kg).

Apart from the "no AECM" grasslands' first mowing yield, which was biased due to a lack of cloud-free images in May, the S2-based DM and N% generally confirm the trends observed on the field measurements and show overall consistent results.

5.4 Discussion

5.4.1 Biophysical variable retrieval

Previous studies on grassland biophysical variable retrieval have shown encouraging results, but were often conducted on small study areas (Ali et al., 2016b; Wang et al., 2019; Cisneros et al., 2020; Chiarito et al., 2021) and/or short periods (Zhang et al., 2015; Quan et al., 2017; Schwieder et al., 2020). In this study, 154 field measurements made during three growing seasons in 59 parcels, spread across three agroecological regions were used to calibrate the retrieval models.

The DM was retrieved with high accuracy during the calibration phase and surprisingly showed even better results on the independent validation dataset. The RMSE are in line with- or lower than those obtained in previous studies (around 1.2 - 1.8 t/ha) on smaller areas (Wang et al., 2019; Cisneros et al., 2020).

Much fewer studies were conducted on N monitoring in grasslands. Recently, Cisneros et al. (2020) used field spectroscopy to simulate S2 bands and retrieved N% on experimental plots with an RMSE of 3.4 g/kg. Pullanagari et al. (2021) retrieved grassland N% with a normalized RMSE of 14% through a convolutional neural network based on field spectroscopy. Our results based on satellite observations corroborate their findings and highlight the performances of S2-based large-scale grassland N monitoring.

While these previous studies retrieved N concentration, grassland N was retrieved both in terms of concentration (i.e. N%) and in terms of content per land area (i.e. CNC) in this study. From a radiative transfer point of view, the latter should be better related to vegetation reflectances as the total quantity of N in the canopy influences the electromagnetic mechanisms. However, our results, and those of previous studies, show that N% can be directly related to S2 reflectances as well. Moreover, N% was estimated more accurately through the direct approach, than indirectly, based on retrieved DM and CNC (Figure 5.6). While CNC was also better estimated directly than indirectly from DM and N%, the selected bands for the CNC regression model included most of the bands respectively selected for DM and N% models. Furthermore, it is striking that the independent validation results matched more with the theoretical expectations, with a significantly lower normalized RMSE for CNC than for N% retrieval.

The stepwise approach allowed to identify the most explanatory S2 reflectance bands and vegetation indices for DM, N%, and CNC. The red edge and NIR reflectances selected for DM respond to change in biophysical quantities (e.g. biomass and LAI). SWIR reflectance is more related to water content, lignin, cellulose, and senescent material (Hank et al., 2019). Other studies on grassland monitoring have successfully related above-ground biomass to these spectral domains and derived indices (Quan et al., 2017; Wang et al., 2019; Schwieder et al., 2020; Cisneros et al., 2020; Chiarito et al., 2021).

Reflectances and indices derived from this part of the spectrum (red edge to SWIR) were also explanatory for N% and CNC. The red edge is a good indicator for chlorophyll content, which is in turn well related to N (Baret et al., 2007; Delloye et al., 2018; Hank et al., 2019). Spectral signatures in the SWIR have been related to crop protein content (Berger et al., 2020a). Proteins carry a large part of vegetation N content, so this could potentially explain the link with SWIR. Other studies have used SWIR reflectances to directly or indirectly retrieve crop N content (Herrmann et al., 2010; Berger et al., 2020b).

While the variables selected for DM retrieval were limited to the red edge, NIR, and SWIR, for N retrieval reflectances and indices in the visible domain were retained in addition (B4 and NDVI for N% and CNC, B2 and B3 in addition for CNC). Reflectances in the visible domain respond well to pigments (including chlorophyll), which can explain the link with N (Hank et al., 2019). Overall, the S2 variables selected by stepwise multiple linear regression for each biophysical variable seem consistent with known radiative transfer mechanisms.

Recent studies have increasingly used RTM for biophysical variable retrieval, as they are expected to offer better transferability for large-scale and multi-year applications and depend less on field measurements (Quan et al., 2017). Secondary variables such as DM and N% cannot be retrieved directly through deterministic approaches. In that case, RTMs are used to retrieve primary variables, such as LAI, chlorophyll content, or protein content, which can then be linked to DM, N%, or CNC (Weiss et al., 2020). Empirical approaches on the other hand allowed us to directly retrieve DM, N%, and CNC from S2 reflectances. Locally, empirical models can outperform RTM for the retrieval of grassland DM (Schwieder et al., 2020). Their main limitation however remains transferability. In this study, thanks to the extended

training dataset, the empirical models for DM, N%, and CNC retrieval showed high performances and good transferability across the years. In addition to the quantitative validation, the large-scale application of the models showed consistent trends in grassland vegetation status and harvested forage yield in the different management practices across the study area. The transferability of these models over larger extents should however be further tested and would probably require additional training data.

Our results highlighted the relationship between grassland biophysical variables and reflectances in the visible, red edge, NIR and SWIR domain and showed the great potential of S2 for DM, N%, and CNC mapping. Like in other optical remote sensing applications, a main limiting factor is the cloud coverage. In this study, cloud cover prevented the use of a large part of field measurements as no clear S2 acquisition could be associated with them. Moreover, in the large-scale application, the forage yield of 90% of the early mowing events could not be estimated and the nitrogen nutrition status of 13% of the parcels in the study area could not be assessed on the 1^{st} of May, due to persistent cloud cover in the spring of 2019.

Finally, although a lot of information can be derived from S2's multi-spectral data, a higher spectral resolution could probably improve the retrieval performance. Several studies using field spectrometry have linked grassland and crop biophysical variables to subtle spectral profile characteristics that can only be derived from hyperspectral data (Ramoelo et al., 2013; Hank et al., 2019; Pullanagari et al., 2021; Rubo and Zinkernagel, 2022). Recent and upcoming spaceborne high-resolution hyperspectral sensors (e.g. PRISMA, EnMAP, CHIME) represent a great potential for grassland monitoring (Hank et al., 2019; Weiss et al., 2020).

5.4.2 Grassland monitoring perspectives

The large field campaign of 2016-2018 was carried out to evaluate harvested forage yield and quality of grasslands with different AECM in comparison to other permanent grasslands. Such an intensive field campaign provides valuable information, but it is costly and time-consuming. We used the measurements made during this campaign to build regression models to retrieve DM and N% based on satellite

imagery. The combination of these biophysical variable retrieval models with the grassland management units delineation and the mowing detections allowed to extrapolate the field-based analysis, spatially and temporally.

First, the large-scale application consistently depicted regional tendencies in grasslands' spring vegetation status. The estimated spring yield of all types of grasslands was significantly higher in Conroz, followed by Fagne-Famenne and then by Ardenne. The N% was also higher in Condroz, although the differences are less significant. More strikingly, the regional boundaries appeared very clearly on the NNI map in Figure 5.7. These differences in spring vegetation status between regions are probably mostly due to the variation in pedo-climatic conditions, and more specifically to temperature, which has an impact on the vegetation development and thereby on the regional management practices. The differences in management practices can in turn have an impact on the early season vegetation status and thereby further amplify regional trends.

The interpretation of the NNI map in particular should be considered with caution. The N dilution curve has been defined primarily for relatively intensive grasslands dominated by Gramineae species. It has also been shown to be valid for more diversified swards (Louarn et al., 2020). However, it is strongly recommended to limit the use of the N dilution curve to the period before flowering (Lemaire and Gastal, 2016), and the timing of the vegetative growth varies between regions and between vegetation species. Locally, the retrieved NNI could, nevertheless, be used to identify outliers that have potentially received excessive fertilization.

Unlike the spring yield, the average first mowing yield of AECM per region did not follow the expected pedo-climatic gradient. This could be due to regional tendencies in management practices, which are themselves influenced by climatic conditions. In Condroz and Fagne-Famenne, the most productive grasslands are likely to be mown early (before 16/06), while the grasslands with AECM may be characterized by lower productivity. In addition, due to the omission of early mowing events by the Sen4CAP mowing detection method, some of these AECM first mowing yields could actually be second mowing yields of grasslands with no AECM, which would also explain the lower DM. In Ardenne, due to the climatic conditions, the vegetation cycle is

slightly delayed compared to the other two regions, which means that most grasslands (more or less productive) are mown relatively late, and are thereby characterized as AECM, which could explain the higher average first mowing yield of AECM there.

The EO-based DM and N% estimations were also used to depict the impacts of management practices on forage yield and quality in the spring and before harvest, allowing to link inputs and outputs in grassland use intensity. As expected, more intensively managed hay meadows (no AECM) consistently showed higher average DM on the 1^{st} of May, compared to the other management practices. These grasslands are more likely to have been fertilized at the start of the season to enhance their forage production in the spring. This could also explain the higher N% retrieved on these parcels. The differences in N% are however less significant due to the decrease in N concentration during vegetation growth. Across the study area, pastures had the lowest average DM, which can be explained by grazing activity at that time of the season.

Overall, the estimated DM, N%, and NNI showed consistent results and allowed to depict the impacts of grassland management practices and of pedo-climatic conditions on forage yield and quality in the spring and before harvest. These large-scale applications showed the potential of S2-based grassland biophysical variable retrieval to contribute to bridging the knowledge gap on the complex relationships between inputs and outputs in grassland use intensity in different agro-pedo-climatic contexts.

5.5 Conclusions

In this chapter, grasslands DM and N content (N% and CNC) were retrieved from S2 spectral bands and indices through stepwise multiple linear regression. For each biophysical variable, the most explanatory variables were selected, and a model was calibrated, using a large field dataset, collected during 3 growing seasons. The models estimated DM, N%, and CNC on an independent validation dataset with normalized RMSE of 9.8%, 15.8%, and 6.7% respectively, showing good temporal transferability. The models were combined with classified management units and mowing detections to assess the impacts of management practices on spring vegetation status and harvested forage yield.

These large-scale applications showed consistent results compared to the field measurements and allowed to highlight differences between agroecological regions. Grassland biophysical variable retrieval by optical remote sensing is promising and should be further developed, as it can provide exhaustive knowledge on the relationship between inputs and outputs in grassland use intensity.

Chapter 6

Conclusions and perspectives

6.1 Remote sensing of grassland use intensity

The objective of this thesis was to measure grassland use intensity over large areas, thanks to optical and microwave remote sensing. The focus was set on managed grasslands (for food production or habitat conservation) with a full vegetation cover and less than 10% woody plant cover. In the four previous chapters, we developed and evaluated methods, based on Sentinel-1, Sentinel-2, and ancillary data, (i) to classify grassland management practices (i.e. grazing and mowing), (ii) to delineate management units, (iii) to detect the timing and frequency of mowing events and (iv) to estimate forage yield and quality (Figure 6.1). Throughout this work, a particular focus was set on reference data quality and statistical soundness of validation. Issues raised in previous research were addressed to improve existing methods and new methods were developed to further characterize grasslands. Despite the challenges inherent to grassland monitoring, the developed methods characterize grasslands exhaustively and with high thematic precision compared to existing datasets.

6.1.1 Specific achievements and contributions

Recently, in the field of grassland use intensity assessment, particular attention has been given to mowing detection (Table 6.1), which was

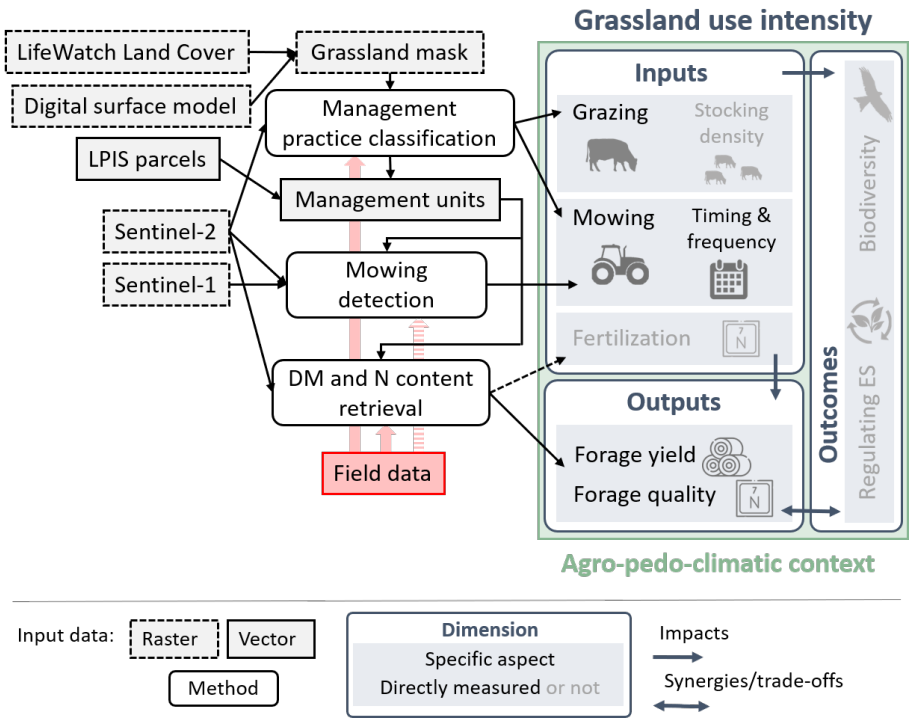

Figure 6.1: Overview of the grassland use intensity (GUI) measurement methods, based on Sentinel-1, Sentinel-2, and ancillary data, developed in the framework of this thesis.

a starting point and a central element in this thesis. Most studies on mowing detection focused on optical imagery. One of the main limitations of optical-based mowing detection is however the omission of mowing events due to cloud cover. Only a few had studied the feasibility of mowing detection with radar data as an alternative, showing encouraging results on small study areas (Schuster et al., 2011; Tamm et al., 2016; Voormansik et al., 2016; Taravat et al., 2019). In chapter 2, we performed the first thorough and comprehensive assessment of the potential and limitations of rule-based mowing detection with S1 radar data. Based on a large and precise field dataset, this study demonstrated that a bit more than half of the mowing events in hay meadows could be detected based on S1 interferometric coherence jumps.

It highlighted the limited precision of S1-based mowing detection and identified grazing as a major confounding factor.

Table 6.1: Overview of the most recent studies on grassland mowing detection methods based on optical and microwave remote sensing.

SATELLITE DATA	DETECTION METHOD	REFERENCE
S1	Machine learning (ANN)	Taravat et al. (2019)
S1	Rule-based	De Vroey et al. (2021)
S2+Landsat-8 (HLS)	Rule-based	Schwieder et al. (2022)
S2	Machine learning (CNN)	Lange et al. (2022)
S1 + HLS	Machine learning (CNN)	Lobert et al. (2021)
S1 + S2	Rule-based	Reinermann et al. (2022)
S1 + S2	Rule-based	De Vroey et al. (2022)

In chapter 3, a multi-source method, combining S1 and S2 mowing detection algorithms, was presented and evaluated. This method, developed as part of the Sen4CAP toolbox (led by our UClouvain research team), was thoroughly assessed in the framework of this thesis. Using a large and comprehensive field dataset, the full potential and limitations of this mowing detection method were quantitatively described in this study. Our analyzes demonstrate the complementarity of S1 and S2 for mowing detection. Moreover, we show the advantages of a rule-based change detection method for large-scale grassland monitoring. Unlike deep learning methods (Lobert et al., 2021), the versatility and transparency of a rule-based method allow to balance between detection rate and precision, depending on the objectives. From a practical perspective, our results show that the confidence level computed with each mowing detection is a good indicator of uncertainty and can be used to filter detections. Finally, beyond the initial framework of the CAP, we showed that this multi-source mowing detection method allows to classify mowing practices and differentiate hay meadows in terms of management intensity. It could therefore be used for large-scale grassland use intensity monitoring.

The conclusions of the first two chapters are corroborated by concurrent studies. Other mowing detection methods, based on optical and/or radar data, achieved similar accuracies in hay meadows. Pastures, however, are either not considered, or identified as a confounding factor to mowing detection. Moreover, most mowing detection methods, especially those based on radar data, are object-based and

highly depend on precise parcel delineation. In chapter 4, these two main issues are tackled. In this study, we develop a new hierarchical approach to differentiate five grassland management practices at the sub-parcel level. First, pastures and hay meadows are differentiated through a pixel-based classification using S2 optical vegetation index time series. We made the hypothesis that mown parcels can be differentiated, based on the occurrence of at least one abrupt change in the time series, which should be absent in pastures. Grazing events with a very high stocking density (e.g. in an intensive rotational grazing system) could cause a similarly abrupt change, and be classified as hay meadows. Nevertheless, since such systems are relatively rare in the study area, our hypothesis is valid for a majority of cases, and we obtained a high classification accuracy. The binary classification was then combined with declared parcel boundaries (LPIS) to delineate spatially homogenous management units for further characterization (i.e. mowing detection). The preliminary detection of grazed parcels and the delineation of homogeneous grassland management units significantly increased the mowing detection performances. This new hierarchical approach, combining a pixel-based classification and an object-based mowing detection method, provided a thematically more accurate and spatially more precise grassland characterization.

The last chapter sets the focus on another dimension of grassland use intensity through biophysical variable retrieval. Some studies had already been conducted on the estimation of grassland dry biomass through optical remote sensing, testing various bands, vegetation indices, and retrieval methods. Very few had explored the retrieval of grassland nitrogen content. In chapter 5, we successfully retrieved grassland dry matter yield, nitrogen content, and nitrogen nutrition index through stepwise multiple linear regression based on S2 reflectances and vegetation indices. Recent studies have increasingly used radiative transfer models for biophysical variable retrieval, as they rely less on training data and are more transferable. Nevertheless, thanks to a large dataset of field measurements from four growing seasons, the empirical models built in this study showed high performances and good temporal transferability. Results showed that nitrogen content could be retrieved accurately both in terms of N concentration (in g/kg dry matter) and canopy nitrogen content (in kg/ha). The explanatory bands and indices selected empirically for each bio-

physical variable through the stepwise approach were consistent with known radiative transfer processes.

In addition, we went beyond the model development and validation and presented two regional applications. Using the EO-based biophysical variables in combination with the grassland classification and the mowing detection method, we evaluated the impacts of management practices on the nitrogen nutrition status and the forage yield for different agroecological regions. This spatially comprehensive analysis showed the potential of biophysical variable retrieval by optical remote sensing to provide exhaustive knowledge on the relationship between inputs and outputs in grassland use intensity.

6.1.2 Generalization potential

Linking all the methods developed in this thesis in a single workflow is possible. It is even recommended because the outputs of one method can greatly improve the performances of the following (Figure 6.1). First, a grassland mask is built, based on the LifeWatch land cover product (Radoux et al., 2023) and a digital surface model. Masked S2 image time series are then used for the pixel-based management practice classification differentiating grazed grasslands (pastures) and mown grasslands (hay meadows and mixed practices). The classification is combined with the parcel boundaries from the regional LPIS to delineate homogeneous grassland management units, which facilitate further characterization. The automated object-based mowing detection method of Sen4CAP, based on S1 and S2, is applied to further differentiate hay meadows by the timing and frequency of mowing events. Finally, S2-based dry matter yield and N content retrieval models are applied to the classified grassland management units to estimate the forage yield of hay meadows, as an output of grassland use intensity. The biophysical variable retrieval models also allow monitoring of early season nitrogen nutrition status, which can be used for N fertilization management.

This workflow could be adapted and generalized to measure GUI across regions and grassland types to facilitate agricultural and ecological monitoring.

The main inputs, i.e. S1 and S2 images, as well as their processing tools (SNAP, sen2cor, MAJA, Fmask, Sen4CAP) are freely avail-

able globally. Beyond satellite imagery, ancillary data and products are needed as well. The LifeWatch land cover layer and regional DSM used to build the grassland mask were a great asset for the pixel-based classification and for the further steps as they allowed to work on only pure grassland pixels without shadows. These specific products cover only Belgium. However, such very high-resolution data are increasingly available at regional and national scales. The LifeWatch land cover product was already extended to Southern Europe at 10 m resolution. Several other countries have freely available high-resolution national LIDAR databases to build DSM. Similar grassland masks could thereby be built for grassland use intensity assessment in other regions. Finally, the parcel boundaries from the regional LPIS dataset were used as a basis to delineate smaller management units. In the framework of the CAP, all EU Member States use the LPIS to record agricultural parcel boundaries and crop types, including grasslands. In regions where parcel boundaries are not recorded or not available, segmentation tools could be applied to retrieve agricultural parcels from satellite imagery (Tetteh et al., 2020).

With some adjustments, the three methods could be reproduced in other regions and seasons. To start with, the most operational one is the Sen4CAP mowing detection method. The algorithms were developed over multiple growing seasons in the framework of the EU CAP reform. The system was designed to support EU paying agencies' compliance assessments, as part of the CAP's new monitoring systems. The exhaustive performance analysis of chapter 3 showed the robustness and versatility of this method. Its operational design allows to adapt detection thresholds and other parameters to specific goals and local conditions, based on expert knowledge or small reference datasets. Furthermore, in chapters 3, 4, and 5, we showed this mowing detection method could be used beyond its initial CAP framework. The multi-source grassland mowing detection method could potentially be used for differentiating mowing dynamics in most hay meadows worldwide and through seasons without much additional reference data. Finally, although it was designed for detecting mechanical mowing events, it could potentially also be used to monitor grazing events in intensive rotational pasture systems as well, since the reduction in biomass would be similar to a mowing event. This would however need to be further tested and

would require preliminary contextual data, to be able to identify such systems and differentiate them from intensive hay meadows.

The pixel-based classification presented in chapter 4 differentiates pastures and hay meadows, based on S2 vegetation index time series. The time series were gap-filled and interpolated on a regular temporal grid, to overcome the multiple-day offset between adjacent satellite tracks, so the classification could be applied over larger areas (Inglada et al., 2015). The classifier trained with the field data collected across Wallonia could potentially differentiate pastures from hay meadows in other regions with similar grassland management and pedo-climatic conditions. However, for more extended applications, new classifiers would need to be trained and tested with adequate reference datasets, as vegetation index profiles of grazed and mown grasslands vary across seasons and regions.

Moreover, the classification was limited to two classes of grassland management, i.e. grazed and mown grasslands. There are however also e.g. grasslands with mixed practices and unmanaged grasslands (i.e. with no anthropogenic activities in the season). Unmanaged grasslands were not considered here, and mixed practices were included in the category of mown grasslands. Overall, to generalize this approach, more classes should be considered.

While the binary classification in itself cannot be directly generalized, the hierarchical grassland characterization approach could be applied, even beyond the scope of grasslands defined in this thesis (cf. section 1.2.3). We showed that such a classification could be used to retrieve homogenous grassland units that can then be further characterized. Initial classes should be defined, depending on the regional context and on the goals of the characterization. Grasslands could, for example, be differentiated in terms of LU (e.g. natural, semi-natural, recreational...) and in terms of pedoclimatic context and vegetation type (e.g. semi-arid steppes, tropical savannas ...).

Finally, the biophysical variable retrieval models were calibrated on field measurements made during 3 growing seasons and the validation in chapter 5 showed a good transferability to yet another season. Furthermore, the regional applications showed coherent results across the study area. The models would, however, probably need to be recalibrated in other regions and other grassland types. The empirical approach allowed to directly retrieve secondary biophysical variables

such as the dry matter yield and the nitrogen concentration, from S2 reflectances, without prior assumptions or knowledge of the complex radiative transfer mechanisms. The main drawback is, however, its limited transferability and dependence on reference data. Deterministic approaches, such as RTM, are generally more suitable for generalization.

Even inside the limited geographic and thematic scope of this thesis, it appeared that multiple qualitative and quantitative indicators are necessary to accurately characterize grasslands and measure GUI. This is expected to be even more crucial at a global scale, considering the full diversity of grasslands. Overall, though all the methods developed in this work are not directly transferrable to any type of grassland, the proposed framework for GUI measurement (Figure 6.1) and the approaches adopted to characterize grasslands remain valid and could be applied for larger scopes.

6.1.3 Further developments

Furthermore, new methods could be developed to measure other aspects of GUI, and other — current or future — satellite sensors with higher spatial, temporal, or spectral resolution could be used to further characterize grasslands.

Over the past years since this thesis was initiated, a number of studies have developed automated grassland mowing detection methods based on optical and/or microwave remote sensing (Taravat et al., 2019; Griffiths et al., 2020; De Vroey et al., 2021; Lobert et al., 2021; Schwieder et al., 2022; Lange et al., 2022; De Vroey et al., 2022; Reinermann et al., 2022). Working on different study areas, using various data combinations and detection methods to achieve the same goal, these studies converge towards similar results and conclusions (Table 6.1). At this stage, it would be interesting to perform a thorough comparative study, combining all the most recent methods and cross-testing them with the reference datasets and validation approaches proposed in this thesis and other studies. This would provide a great overview and baseline for further research on grassland monitoring.

In addition to mowing, other management practices could be detected and monitored by remote sensing. In chapter 4, pastures were

differentiated from hay meadows based on VI time series classification. Grazing activities were however not detected or further characterized. Grazing intensity patterns are often derived from livestock census data (Gómez Giménez et al., 2017; Estel et al., 2018; Piipponen et al., 2022) or vegetation indices linked to above-ground biomass, such as the NDVI (Reeves and Baggett, 2014; Ma et al., 2019; Xu et al., 2019). Further research should be carried out on grazing intensity assessment, especially in intensive mixed regimes of grazing and mowing where both activities are challenging to differentiate.

Some more intensively managed grasslands are regularly tilled and resown. This provides information on the continuity of a grassland cover, which has an impact on the ecological value of grasslands (Kuhn et al., 2021). Tillage could be detected, similarly to mowing events, based on changes in S1 coherence and S2 NDVI time series (Voormansik et al., 2020). In some regions, monitoring grassland irrigation is crucial in the context of water management. Bazzi et al. (2020) successfully detected irrigated grasslands based on S1 VV backscattering.

For the biophysical characterization of grasslands, the focus was set on the retrieval of biomass and N content. Other macronutrients, such as phosphorus, can however provide key information on grasslands' ecological state and productivity. As a (co-)limiting element in plant growth, phosphorus availability is determinant for grassland species diversity (Ceulemans et al., 2017). Moreover, phosphorus constitutes a crucial nutrient for herbivores and influences the grazing behavior of wildlife and livestock (Gao et al., 2019). A few studies have been performed on crop or grassland phosphorus content retrieval through field spectroscopy (Mutanga and Kumar, 2007; Ramoelo et al., 2011; Pimstein et al., 2011; Mahajan et al., 2017; Gao et al., 2019). Their results show significant correlations in the NIR and SWIR region of the spectrum and suggest plant phosphorus content could be retrieved from multispectral sensors such as S2. Research should further investigate the potential retrieval of grassland phosphorus as well as other nutrients (e.g. potassium) from multi- and hyperspectral satellite imagery.

Alternative satellite data constitute another potential for improvements and developments. In chapter 3, the complementarity of S1 and S2 for grassland mowing detection was demonstrated. While microwave and optical remote sensing measure very different aspects of

the soil and vegetation cover, they both allow to detect mowing events, through changes in vegetation structure and humidity and in green biomass respectively. Where and when enough cloud-free images are available, S2-based mowing detection is more precise and accurate. S1-based mowing detection partially compensates for omissions due to cloud cover but is characterized by a higher uncertainty. In the near future, daily 5 m resolution multispectral optical data, like the Earth-Daily constellation planned for 2024, should increase detection rates and allow to rely less on microwave data for mowing detection. SAR data will however remain a great asset for the detection of mowing events under persistent cloud cover.

Shorter revisit times, in general, would allow to monitor grasslands growth and management more precisely. For example, based on daily acquisitions, grazing practices could potentially be further characterized. Grazing periods and their intensity might thereby be estimated.

Although the S2's multispectral sensor already allows to retrieve a lot of information, higher spectral resolutions would be an asset, especially for biophysical variable retrieval. Recent and upcoming spaceborne high-resolution hyperspectral sensors (e.g. PRISMA, EnMAP, CHIME) represent a great potential for agricultural monitoring (Hank et al., 2019; Weiss et al., 2020).

Finally, the spatial resolution of S1 and S2 (10-20m) is sufficient to characterize most grasslands in European landscapes. Parcels or management units that are smaller than 1000 m^2 or narrow grasslands are however more challenging to characterize. This can be an issue in particular for remnants of natural and semi-natural grasslands that are often made up of relatively small patches in complex landscapes. Therefore, the use of higher-resolution satellite data should be further considered.

6.2 Perspectives and recommendations

6.2.1 Diversity and complexity of grassland ecosystems

Grasslands encompass a broad range of land covers, land uses, and land use intensities. Both from an ecological and an agricultural perspective, grasslands are diverse. Such diversity necessitates quantitative

indicators for grassland characterization and makes classification challenging. Furthermore, due to the global diversity of grasslands, models can rarely be generalized beyond a given scope of grassland types. It is therefore also crucial to clearly define this scope.

First, grasslands occur in various biotopes (i.e. the abiotic component of an ecosystem). These initial conditions largely determine the state, productivity, and composition of grasslands. Even within a relatively small region, variations in temperature, soil type, and humidity can, for example, have significant impacts on grasslands' biophysical variables, as illustrated in Figure 5.7.

Secondly, grassland management practices are diverse and can be challenging to categorize. Multiple quantitative indicators seem to be most suitable to characterize grasslands in terms of management, input, and output intensity. In this thesis, we found that a combination of qualitative and quantitative indicators is more suitable for characterizing managed grassland. The mowing detection methods allowed to quantitatively characterize grasslands in terms of timing and frequency of mowing events. And productivity could be estimated with relatively high precision and accuracy through biomass and N content retrieval. In terms of grazing, however, the method developed in this thesis only allowed to differentiate pastures, i.e. exclusively grazed grasslands, from hay meadows and mixed practices. Further research should focus on retrieving precise grazing periods and estimating stocking density to further characterize pastures and mixed practices.

Finally, various grassland habitats with characteristic species compositions result from specific biotopes and management practices. Grassland habitat classification systems exist (e.g. EUNIS) and are essential to facilitate field inventories, monitoring, and decision-making. However, due to their diversity, classifying grasslands into defined categories is not recommended because classes are likely to be ambiguous or fuzzy.

Overall, the diversity and complexity of grasslands need to be taken into account for the development and especially the validation of robust monitoring methods. Therefore, field observations and field measurements also play an essential role.

6.2.2 Importance of field campaigns

Comprehensive field data were key for the development and evaluation of the grassland use intensity assessment methods in this thesis. Generally speaking, although remote sensing is a great tool for large-scale LC, LU, and LUI monitoring — as it is far less costly and time-consuming than *in situ* observations and measurements — high-quality reference data remain essential and cannot be overlooked.

On one hand, the performances of empirical methods, such as regressions and supervised machine learning models, greatly depend on the size and quality of the training dataset. On the other, precise, accurate, and representative reference data are essential to perform statistically sound validations and performance analyzes of any remote sensing mapping or monitoring method.

The source of the reference data and the amount of time and energy put into building it should be adapted to meet the required thematic, spatial and temporal precision and range of a specific application. Large reference datasets can be rapidly gathered through expert-based image interpretation. Recently developed tools, such as Collect Earth, facilitate the collection of large reference datasets, based on very high-resolution imagery, for large-scale LC and LU applications (Bey et al., 2016).

While very efficient and essential for, e.g., global LC mapping, image interpretation is more limited thematically and temporally and implies a large uncertainty. For example, in Chapter 3, Planet image interpretation allowed to evaluate the mowing detection method in six countries, but grazing and mowing events could not be differentiated and some mowing events were omitted because of missing Planet images due to cloud cover. The field dataset, on the other hand, allowed to validate the detection of actual mowing events more rigorously, with high certainty and temporal precision. Moreover, validation by image interpretation implies a risk of non-independence between the reference data and the classification result (Radoux and Bogaert, 2020).

Crowdsourcing, i.e. "using contributions from crowds to solve specific problems or to collect data needed for the solution", has been used increasingly in the remote sensing community (Saralioglu and Gungor, 2020). For agricultural monitoring, in particular, farmer surveys and declarations constitute a potential reference data source. The national

LPIS datasets were of great use for the development and initial evaluation of the Sen4CAP mowing detection method. This kind of dataset is, however, not always freely available, often subject to interpretation and lacks consistency over large areas. One of the main barriers to crowdsourcing in the agricultural sector is related to privacy issues, decreasing farmers' willingness to participate in this kind of project. Nevertheless, improving regulations for the use of private data, and developing technological and collaborative farmsourcing approaches represent great perspectives for agricultural reference data collection (Minet et al., 2017). Although crowdsourcing is an effective solution to gather large datasets, it needs to be considered with caution. The quality of the protocol and the adequate use of the data are key.

Besides the source of a reference dataset, its representativeness is also crucial. It is important to consider the size and the thematic, spatial, and temporal range and precision of a reference dataset and balance it against the time and cost of collecting it.

For GUI measurement, and more specifically for management practices characterization, the temporal aspect was essential. The field campaign carried out in the spring and summer of 2019 provided precise information on the management practices of more than 400 permanent grasslands across three agroecological regions of Wallonia. While the field campaign was intensive and time-consuming, this unique reference dataset proved to be highly valuable and served as a strong basis for the developments and analyses in chapters 2, 3 and 4. The same grasslands were visited repeatedly, to record all mowing and grazing practices during a given period. This allowed us to validate the mowing detection methods with a relatively strict approach, reflecting the capacity of detecting the occurrence and the precise timing of each mowing event. The spatial extent and the diversity of monitored parcels allowed us to develop robust grassland mowing detection and classification methods and evaluate their performances in a statistically sound manner. Moreover, through this regular field monitoring, the large variability of management practices within parcels was highlighted.

In Chapter 5, the majority of the reference dataset was based on field measurements made outside the framework of this thesis. Only half of the measurements could be used to calibrate and validate the biophysical variable retrieval models. Nevertheless, since the measurements were made on grasslands spread across Wallonia and at different

stages during three growing seasons, the models showed high performances and provided consistent results across regions and seasons.

In addition to — and depending on — the size and quality of the reference dataset, method performances can be evaluated at different levels, with more or less conservative approaches and using various performance metrics. As discussed in chapter 3, the mowing detection methods were evaluated through a conservative approach, on one hand, reflecting the ability to exactly detect the occurrence and timing of each mowing event. On the other, a parcel-based evaluation was applied, assessing the ability of the method to classify grasslands in terms of mowing practices. In chapter 4, a wall-to-wall pixel-based validation was applied to assess the accuracy of the classification as well as the mowing detection. This allowed to take the size of the parcels into account, which can partially explain the higher mowing detection performance metrics obtained in that chapter. Finally, the DM and N content retrieval models developed in chapter 5 were validated quantitatively and qualitatively. The regional applications, to retrieve forage yield and nitrogen nutrition levels, showed the consistency of the estimated biophysical variables across regions and seasons. In general, we argue that a combination of different levels of performance evaluation provides the best and most complete estimation of a method's potential and limitations and allows to compare results between studies.

Overall, despite the increasing potential of satellite remote sensing for land monitoring, *in situ* data collection remains essential in many EO applications. While they are more costly and time-consuming, field campaigns allow building more precise, accurate, and comprehensive reference datasets for land monitoring. *In situ* observations and measurements offer a higher certainty and independence for the calibration and validation of robust agricultural monitoring methods. Finally, beyond the practical aspect of collecting reference data, field campaigns provide a better insight into the complex reality of LC, LU, and LUI.

6.2.3 Outcomes of grassland use intensity

It is broadly established that grassland management affects the delivery of provisioning and regulating ES (Sollenberger et al., 2019) as well as biodiversity. The outcomes of GUI (i.e. impacts on regulating ES and biodiversity) were not addressed directly in this thesis. The focus was

set primarily on inputs, mostly through the characterization of management practices, and on outputs, through forage yield estimation. Given the interconnections that exist between different aspects of GUI, precise and accurate information on one aspect alone can however already be extremely valuable and allow to deduct other characteristics.

For instance, grasslands that are mown at least three times per season, with an early first mowing event, are often fertilized to boost the forage yield. Such intensively manage hay meadows will often be species-poor as the fertilization and the frequent cutting benefits only a few dominant species. Knowing the uncertainties inherent to the detection, classification and retrieval methods, statistics of grassland management practices and yield can be computed, and regional tendencies can be analyzed. For example, spatially explicit information on grassland management practices could be linked to species occurrence data to further study the link with biodiversity. Satellite-based grassland characterization could also be used to target specific types of grasslands or stratify samplings for field inventories. And the regional forage yield retrieval could perhaps be used in the context of studies on the protein self-sufficiency of livestock farming.

While current grassland management and productivity play a significant role, they are not sufficient to explain the whole variability in habitats and ecological states of grasslands. The impacts of given management practices on biodiversity and regulating ES vary in function of (i) biotopes, (ii) grassland continuity, and management history. A biotope is defined as an area with relatively uniform abiotic (i.e. geographical, (micro)climatic, pedological, geological, hydrological, and topographical) conditions. Grasslands occur in the natural succession of certain biotopes, but in most temperate regions of Europe, grasslands result from — and are maintained by — anthropogenic activities. The current state of grasslands greatly depends on initial conditions and on long-term LU and LUI history, which will also influence the impacts of present management practices on biodiversity and ES (Ahlering et al., 2019; Janišová et al., 2021; Kuhn et al., 2021).

Satellite-based grassland input and output intensity measurements could be combined with existing pedological, topographical, hydrological, and climatic data to include biotopes in the characterization. For grassland continuity and management history, satellite data allow going back in time only to a limited extent. Other data sources, such as

historical maps and cadastral and agricultural records (e.g. LPIS), can be used to reconstruct LU and LUI history (Janišová et al., 2021; Kuhn et al., 2021). Crossing biotope, historical LU and current GUI datasets could, for example, be useful to stratify samplings for field inventories, to relate GUI to species occurrences and further study the link with biodiversity, or to highlight underrepresented habitats in agri-environmental schemes and better target them. Overall, such a comprehensive dataset would provide a solid basis to analyze grassland ecosystems and guide management and conservation plans.

6.3 Final thoughts

As one of the main drivers of global change, the agriculture, forestry, and other land use sector also constitutes a major part of the solution to mitigate global change. It is, now more than ever, urgent to develop and implement sustainable land use solutions, that take into account local and global cross-sector synergies and trade-offs, and mitigate climate change, while contributing to biodiversity conservation, ecosystem functioning, and many other sustainable development goals. Comprehensive data on land cover, land use, and land use intensity are necessary to address the many knowledge gaps in the development of land use solutions. In their 6^{th} report, the IPCC recommends that research priorities include, inter alia, "improved (real-time and cheap) measurement, reporting and verification" (IPCC, 2022). Moreover, the COP27 has stressed the "vital importance" of Earth observation systems to further understand climate change, to support adaptation and mitigation plans, and more specifically to develop early warning systems.

Satellite remote sensing is a great tool to exhaustively measure, map, and monitor land cover, land use, and land use intensity. Recent satellites, combining global coverage with high spatial and temporal resolution enabled tremendous progress in Earth observation. In this thesis, we measured several aspects of grassland use intensity at the sub-parcel level, over large areas, and long periods, using Sentinel-1 and Sentinel-2. These measurements could be implemented in large-scale agricultural and ecological monitoring and, hopefully, contribute to supporting the development and implementation of sustainable land use solutions.

Bibliography

Abdalla, M., Hastings, A., Chadwick, D., Jones, D.L., Evans, C., Jones, M.B., Rees, R., Smith, P., 2018. Critical review of the impacts of grazing intensity on soil organic carbon storage and other soil quality indicators in extensively managed grasslands. Agriculture, Ecosystems & Environment 253, 62–81.

Adjorlolo, C., Mutanga, O., Cho, M.A., 2014. Estimation of canopy nitrogen concentration across c3 and c4 grasslands using worldview-2 multispectral data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 4385–4392.

Ahlering, M.A., Johnson, D.H., Elliott, L.H., 2019. Land ownership and use influence grassland bird abundance. The Journal of Wildlife Management 83, 343–355.

Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., Green, S., 2016a. Satellite remote sensing of grasslands: from observation to management. Journal of Plant Ecology 9, 649–671.

Ali, I., Cawkwell, F., Dwyer, E., Green, S., 2016b. Modeling managed grassland biomass estimation by using multitemporal remote sensing data—a machine learning approach. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10, 3254–3264.

Allan, E., Bossdorf, O., Dormann, C.F., Prati, D., Gossner, M.M., Tscharntke, T., Blüthgen, N., Bellach, M., Birkhofer, K., Boch, S., et al., 2014. Interannual variation in land-use intensity enhances grassland multidiversity. Proceedings of the National Academy of Sciences 111, 308–313.

Arino, O., Ramos Perez, J., Kalogirou, V., Van Bogaert, E., Defourny, P., Bontemps, S., 2012. Global land cover map for 2009 (GlobCover 2009) .

Asam, S., Klein, D., Dech, S., 2015. Estimation of grassland use intensities based on high spatial resolution LAI time series. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-7/W3, 285–291.

Baetens, L., Desjardins, C., Hagolle, O., 2019. Validation of copernicus Sentinel-2 cloud masks obtained from Maja, Sen2Cor, and FMask processors using reference cloud masks generated with a supervised active learning procedure. Remote Sensing 11, 433.

Bai, Y., Cotrufo, M.F., 2022. Grassland soil carbon sequestration: Current understanding, challenges, and solutions. Science 377, 603–608.

Baret, F., Houlès, V., Guerif, M., 2007. Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. Journal of experimental botany 58, 869–880.

Bargiel, D., Herrmann, S., Lohmann, P., Sörgel, U., 2010. Land use classification with high-resolution satellite radar for estimating the impacts of land use change on the quality of ecosystem services. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences: [100 Years ISPRS Advancing Remote Sensing Science, Pt 1] 38 (2010), Nr. 7B 38, 68–73. doi:http://dx.doi.org/10.15488/1116.

Barrett, B., Nitze, I., Green, S., Cawkwell, F., 2014. Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in ireland using machine learning approaches. Remote sensing of environment 152, 109–124.

Barrett, B.W., Dwyer, E., Whelan, P., 2009. Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques. Remote Sensing 1, 210–242.

Bazzi, H., Baghdadi, N., Fayad, I., Charron, F., Zribi, M., Belhouchette, H., 2020. Irrigation events detection over intensively irrigated grassland plots using sentinel-1 data. Remote Sensing 12, 4058.

Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote sensing of environment 204, 509–523.

Bengtsson, J., Bullock, J., Egoh, B., Everson, C., Everson, T., O'Connor, T., O'Farrell, P., Smith, H., Lindborg, R., 2019. Grasslands—more important for ecosystem services than you might think. Ecosphere 10, e02582.

Berger, K., Verrelst, J., Féret, J.B., Hank, T., Wocher, M., Mauser, W., Camps-Valls, G., 2020a. Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. International Journal of Applied Earth Observation and Geoinformation 92, 102174.

Berger, K., Verrelst, J., Feret, J.B., Wang, Z., Wocher, M., Strathmann, M., Danner, M., Mauser, W., Hank, T., 2020b. Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. Remote Sensing of Environment 242, 111758.

Bey, A., Sánchez-Paus Díaz, A., Maniatis, D., Marchi, G., Mollicone, D., Ricci, S., Bastin, J.F., Moore, R., Federici, S., Rezende, M., et al., 2016. Collect earth: Land use and land cover assessment through augmented visual interpretation. Remote Sensing 8, 807.

Blaes, X., Defourny, P., Derauw, D., Barbier, C., 1999. InSAR coherence for crop parameter monitoring, in: Proc. FRINGE, Citeseer. pp. 10–12.

Blüthgen, N., Dormann, C.F., Prati, D., Klaus, V.H., Kleinebecker, T., Hölzel, N., Alt, F., Boch, S., Gockel, S., Hemp, A., et al., 2012. A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization. Basic and Applied Ecology 13, 207–220.

Boegh, E., Houborg, R., Bienkowski, J., Braban, C.F., Dalgaard, T., Van Dijk, N., Dragosits, U., Holmes, E., Magliulo, V., Schelde, K., et al., 2013. Remote sensing of lai, chlorophyll and leaf nitrogen pools of crop-and grasslands in five european landscapes. Biogeosciences 10, 6279–6307.

Bontemps, S., Bajec, K., Cara, C., Defourny, P., De Vendictis, L., Heymans, D., Kucera, L., Malcorps, P., Milcinski, G., Nicola, L., et al., 2020. Sen4cap—sentinels for common agricultural policy. System Software User Manual. Sen4CAP_SUM_v1 2.

Bontemps, S., de Vendictis, L., Malcorps, P., Sciarretta, C., Zavagli, M., 2021. Sen4CAP - Sentinels for Common Agriculture Policy - Validation report.

Boughorbel, S., Jarray, F., El-Anbari, M., 2017. Optimal classifier for imbalanced data using matthews correlation coefficient metric. PloS one 12, e0177678.

Brovkin, V., Sitch, S., Von Bloh, W., Claussen, M., Bauer, E., Cramer, W., 2004. Role of land cover changes for atmospheric co2 increase and

climate change during the last 150 years. Global Change Biology 10, 1253–1266.

Busch, V., Klaus, V.H., Schäfer, D., Prati, D., Boch, S., Müller, J., Chisté, M., Mody, K., Blüthgen, N., Fischer, M., et al., 2019. Will i stay or will i go? plant species-specific response and tolerance to high land-use intensity in temperate grassland ecosystems. Journal of vegetation science 30, 674–686.

Butchart, S.H., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J.P., Almond, R.E., Baillie, J.E., Bomhard, B., Brown, C., Bruno, J., et al., 2010. Global biodiversity: indicators of recent declines. Science 328, 1164–1168.

Ceulemans, T., Bodé, S., Bollyn, J., Harpole, S., Coorevits, K., Peeters, G., Van Acker, K., Smolders, E., Boeckx, P., Honnay, O., 2017. Phosphorus resource partitioning shapes phosphorus acquisition and plant species abundance in grasslands. Nature Plants 3, 1–7.

Chang, J., Ciais, P., Gasser, T., Smith, P., Herrero, M., Havlík, P., Obersteiner, M., Guenet, B., Goll, D.S., Li, W., et al., 2021. Climate warming from managed grasslands cancels the cooling effect of carbon sinks in sparsely grazed and natural grasslands. Nature Communications 12, 1–10.

Chen, J., Chen, J., 2018. Globeland30: Operational global land cover mapping and big-data analysis. Science China. Earth Sciences 61, 1533–1534.

Chiarito, E., Cigna, F., Cuozzo, G., Fontanelli, G., Mejia Aguilar, A., Paloscia, S., Rossi, M., Santi, E., Tapete, D., Notarnicola, C., 2021. Biomass retrieval based on genetic algorithm feature selection and support vector regression in alpine grassland using ground-based hyperspectral and sentinel-1 sar data. European Journal of Remote Sensing 54, 209–225.

Chisté, M.N., Mody, K., Gossner, M.M., Simons, N.K., Köhler, G., Weisser, W.W., Blüthgen, N., 2016. Losers, winners, and opportunists: How grassland land-use intensity affects orthopteran communities. Ecosphere 7, e01545.

Chytrỳ, M., Tichỳ, L., Hennekens, S.M., Knollová, I., Janssen, J.A., Rodwell, J.S., Peterka, T., Marcenò, C., Landucci, F., Danihelka, J., et al., 2020. Eunis habitat classification: Expert system, characteristic species combinations and distribution maps of european habitats. Applied Vegetation Science 23, 648–675.

Cisneros, A., Fiorio, P., Menezes, P., Pasqualotto, N., Van Wittenberghe, S., Bayma, G., Furlan Nogueira, S., 2020. Mapping productivity and essential biophysical parameters of cultivated tropical grasslands from sentinel-2 imagery. Agronomy 10, 711.

Clevers, J.G., Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on sentinel-2 and-3. International Journal of Applied Earth Observation and Geoinformation 23, 344–351.

Curnel, Y., 2015. Satellite remote sensing priorities for better assimilation in crop growth models : winter wheat LAI and grassland mowing dates case studies.

De Vroey, M., de Vendictis, L., Zavagli, M., Bontemps, S., Heymans, D., Radoux, J., Koetz, B., Defourny, P., 2022. Mowing detection using sentinel-1 and sentinel-2 time series for large scale grassland monitoring. Remote Sensing of Environment 280, 113145. doi:https://doi.org/10.1016/j.rse.2022.113145.

De Vroey, M., Radoux, J., Defourny, P., 2021. Grassland mowing detection using Sentinel-1 time series: Potential and limitations. Remote Sensing 13, 348.

De Zan, F., Parizzi, A., Prats-Iraola, P., López-Dekker, P., 2013. A SAR interferometric model for soil moisture. IEEE Transactions on Geoscience and Remote Sensing 52, 418–425.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T., et al., 2019. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. Remote sensing of environment 221, 551–568.

Defourny, P., Bontemps, S., Brockman, C., Herold, M., Kirches, G., Lamarche, C., Kalogirou, V., Seifert, F., Arino, O., 2013. CCI and the land cover_cci project. 3rd GOFC-GOLD Symposium.

Delloye, C., Weiss, M., Defourny, P., 2018. Retrieval of the canopy chlorophyll content from sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems 216, 245–261. doi:10.1016/j.rse.2018.06.037.

DEMNA, 2010. Les prairies et les pelouses (E) - biotopes - la Biodiversité en Wallonie. http://biodiversite.wallonie.be/fr/prairies-et-pelouses-e.html?IDC=815. Accessed: 01.12.2020.

Dengler, J., Biurrun, I., Boch, S., Dembicz, I., Török, P., 2020. Grasslands of the palaearctic biogeographic realm: introduction and synthesis. Encyclopedia of the world's biomes 3, 617–637.

Dengler, J., Janišová, M., Török, P., Wellstein, C., 2014. Biodiversity of palaearctic grasslands: a synthesis. Agriculture, Ecosystems & Environment 182, 1–14.

Di Gregorio, A., 2005. Land cover classification system: classification concepts and user manual: LCCS. volume 2. Food & Agriculture Org.

Dixon, A., Faber-Langendoen, D., Josse, C., Morrison, J., Loucks, C., 2014. Distribution mapping of world grassland types. Journal of biogeography 41, 2003–2019.

Dufrêne, M., Delescaille, L.M., 2005. La typologie waleunis des biotopes wallons .

Duraiappah, A.K., Naeem, S., Agardy, T., Ash, N.J., Cooper, H.D., Díaz, S., Faith, D.P., Mace, G., McNeely, J.A., Mooney, H.A., et al., 2005. Ecosystems and human well-being: biodiversity synthesis; a report of the millennium ecosystem assessment .

Dusseux, P., Hubert-Moy, L., Corpetti, T., Vertès, F., 2015. Evaluation of spot imagery for the estimation of grassland biomass. International Journal of Applied Earth Observation and Geoinformation 38, 72–77.

Dusseux, P., Vertès, F., Corpetti, T., Corgne, S., Hubert-Moy, L., 2014. Agricultural practices in grasslands detected by spatial remote sensing. Environmental Monitoring and Assessment 186, 8249–8265. doi:10.1007/s10661-014-4001-5.

d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H.I., Joebges, C., Lemoine, G., et al., 2020. Harmonised lucas in-situ land cover and use database for field surveys from 2006 to 2018 in the european union. Scientific data 7, 1–15.

Erb, K.H., Haberl, H., Jepsen, M.R., Kuemmerle, T., Lindner, M., Müller, D., Verburg, P.H., Reenberg, A., 2013. A conceptual framework for analysing and measuring land-use intensity. Current opinion in environmental sustainability 5, 464–470.

Erb, K.H., Kastner, T., Plutzar, C., Bais, A.L.S., Carvalhais, N., Fetzel, T., Gingrich, S., Haberl, H., Lauk, C., Niedertscheider, M., et al., 2018. Unexpectedly large impact of forest management and grazing on global vegetation biomass. Nature 553, 73–76.

Erb, K.H., Luyssaert, S., Meyfroidt, P., Pongratz, J., Don, A., Kloster, S., Kuemmerle, T., Fetzel, T., Fuchs, R., Herold, M., et al., 2017. Land management: data availability and process understanding for global change studies. Global change biology 23, 512–533.

ESA, v6.0. Sentinel application platform (snap). http://step.esa.int.

Estel, S., Mader, S., Levers, C., Verburg, P.H., Baumann, M., Kuemmerle, T., 2018. Combining satellite data and agricultural statistics to map grassland management intensity in Europe. Environmental Research Letters 13, 074020.

Farinelle, A., 2020. Projet : Fourrages natura 2000 et maec (fourrages mieux). https://www.fourragesmieux.be/projet.html. Accessed: 2022-10-10.

Fazzini, P., De Felice Proia, G., Adamo, M., Blonda, P., Petracchini, F., Forte, L., Tarantino, C., 2021. Sentinel-2 remote sensed image classification with patchwise trained convnets for grassland habitat discrimination. Remote Sensing 13, 2276.

Felipe-Lucia, M.R., Soliveres, S., Penone, C., Fischer, M., Ammer, C., Boch, S., Boeddinghaus, R.S., Bonkowski, M., Buscot, F., Fiore-Donno, A.M., et al., 2020. Land-use intensity alters networks between biodiversity, ecosystem functions, and services. Proceedings of the National Academy of Sciences 117, 28140–28149.

Filella, I., Penuelas, J., 1994. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. International journal of remote sensing 15, 1459–1470.

Franke, J., Keuck, V., Siegert, F., 2012. Assessment of grassland use intensity by remote sensing to support conservation schemes. Journal for Nature Conservation 20, 125–134.

Gao, J., Meng, B., Liang, T., Feng, Q., Ge, J., Yin, J., Wu, C., Cui, X., Hou, M., Liu, J., et al., 2019. Modeling alpine grassland forage phosphorus based on hyperspectral remote sensing and a multi-factor machine learning algorithm in the east of tibetan plateau, china. ISPRS Journal of Photogrammetry and Remote Sensing 147, 104–117.

Garioud, A., Valero, S., Giordano, S., Mallet, C., 2021. Recurrent-based regression of Sentinel time series for continuous vegetation monitoring. Remote Sensing of Environment 263, 112419.

Gastal, F., Lemaire, G., 2002. N uptake and distribution in crops: an agronomical and ecophysiological perspective. Journal of experimental botany 53, 789–799.

Griffiths, P., Nendel, C., Pickert, J., Hostert, P., 2020. Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and Landsat time series. Remote Sensing of Environment 238, 111124.

Grigulis, K., Lavorel, S., Krainer, U., Legay, N., Baxendale, C., Dumont, M., Kastl, E., Arnoldi, C., Bardgett, R.D., Poly, F., et al., 2013. Relative contributions of plant traits and soil microbial properties to mountain grassland ecosystem services. Journal of Ecology 101, 47–57.

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., Cresson, R., 2017. Orfeo toolbox: open source processing of remote sensing images. Open Geospatial Data, Software and Standards 2, 1–8.

Gómez Giménez, M., de Jong, R., Della Peruta, R., Keller, A., Schaepman, M.E., 2017. Determination of grassland use intensity based on multi-temporal remote sensing data and ecological indicators. Remote Sensing of Environment 198, 126–139.

Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, Landsat and Sentinel-2 images. Remote Sensing of Environment 114, 1747–1755.

Hank, T.B., Berger, K., Bach, H., Clevers, J.G., Gitelson, A., Zarco-Tejada, P., Mauser, W., 2019. Spaceborne imaging spectroscopy for sustainable agriculture: Contributions and challenges. Surveys in Geophysics 40, 515–551.

Hansen, M.C., Defries, R.S., Townshend, J.R.G., Sohlberg, R., 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. International Journal of Remote Sensing 21, 1331–1364. doi:10.1080/014311600210209.

Hardy, T., Kooistra, L., Domingues Franceschini, M., Richter, S., Vonk, E., van den Eertwegh, G., van Deijl, D., 2021. Sen2Grass: A cloud-based solution to generate field-specific grassland information derived from Sentinel-2 imagery. AgriEngineering 3, 118–137.

He, W., Yokoya, N., 2018. Multi-temporal Sentinel-1 and-2 data fusion for optical image simulation. ISPRS International Journal of Geo-Information 7, 389.

Herrero, M., Havlík, P., Valin, H., Notenbaert, A., Rufino, M.C., Thornton, P.K., Blümmel, M., Weiss, F., Grace, D., Obersteiner, M., 2013. Biomass use, production, feed efficiencies, and greenhouse gas emis-

sions from global livestock systems. Proceedings of the National Academy of Sciences 110, 20888–20893.

Herrero, M., Henderson, B., Havlík, P., Thornton, P.K., Conant, R.T., Smith, P., Wirsenius, S., Hristov, A.N., Gerber, P., Gill, M., et al., 2016. Greenhouse gas mitigation potentials in the livestock sector. Nature Climate Change 6, 452–461.

Herrmann, I., Karnieli, A., Bonfil, D., Cohen, Y., Alchanatis, V., 2010. Swir-based spectral indices for assessing nitrogen content in potato fields. International Journal of Remote Sensing 31, 5127–5143.

Horn, D.J., Koford, R.R., 2000. Relation of grassland bird abundance to mowing of conservation reserve program fields in north dakota. Wildlife Society Bulletin (1973-2006) 28, 653–659.

Howison, R.A., Piersma, T., Kentie, R., Hooijmeijer, J.C.E.W., Olff, H., 2018. Quantifying landscape-level land-use intensity patterns through radar-based remote sensing. Journal of Applied Ecology 55, 1276–1287. doi:10.1111/1365-2664.13077.

Hudewenz, A., Klein, A.M., Scherber, C., Stanke, L., Tscharntke, T., Vogel, A., Weigelt, A., Weisser, W.W., Ebeling, A., 2012. Herbivore and pollinator responses to grassland management intensity along experimental changes in plant species richness. Biological Conservation 150, 42–52.

Humbert, J.Y., Pellet, J., Buri, P., Arlettaz, R., 2012. Does delaying the first mowing date benefit biodiversity in meadowland? Environmental Evidence 1, 1–13.

Inglada, J., 2016. Otb gapfilling, a temporal gapfilling for image time series library. Disponible sur: http://tully. ups-tlse. fr/jordi/temporalgapfilling .

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., et al., 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. Remote Sensing 7, 12356–12379.

IPCC, 2021. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. volume In Press. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. doi:10.1017/9781009157896.

IPCC, 2022. 'Summary for Policymakers' in Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. volume In Press. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Isbell, F., Balvanera, P., Mori, A.S., He, J.S., Bullock, J.M., Regmi, G.R., Seabloom, E.W., Ferrier, S., Sala, O.E., Guerrero-Ramírez, N.R., et al., 2022. Expert perspectives on global biodiversity loss and its drivers and impacts on people. Frontiers in Ecology and the Environment .

Jacob, A.W., Vicente-Guijalba, F., Lopez-Martinez, C., Lopez-Sanchez, J.M., Litzinger, M., Kristen, H., Mestre-Quereda, A., Ziółkowski, D., Lavalle, M., Notarnicola, C., et al., 2020. Sentinel-1 InSAR coherence for land cover mapping: A comparison of multiple feature-based classifiers. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 535–552.

Janišová, M., Iuga, A., Ivașcu, C.M., Magnes, M., 2021. Grassland with tradition: sampling across several scientific disciplines. Vegetation Classification and Survey 2, 19–35.

Johansen, L., Westin, A., Wehn, S., Iuga, A., Ivascu, C.M., Kallioniemi, E., Lennartsson, T., 2019. Traditional semi-natural grassland management with heterogeneous mowing times enhances flower resources for pollinators in agricultural landscapes. Global Ecology and Conservation 18, e00619.

Johns Hopkins University, 2018. ECOSTRESS spectral library. http://https://speclib.jpl.nasa.gov/.

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M.R., Kuemmerle, T., Meyfroidt, P., Mitchard, E.T., et al., 2016. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. Remote Sensing 8, 70.

Kaasiku, T., Praks, J., Jakobson, K., Rannap, R., 2021. Radar remote sensing as a novel tool to assess the performance of an agri-environment scheme in coastal grasslands. Basic and Applied Ecology 56, 464–475.

Kayad, A., Rodrigues Jr, F.A., Naranjo, S., Sozzi, M., Pirotti, F., Marinello, F., Schulthess, U., Defourny, P., Gerard, B., Weiss, M., 2022. Radiative transfer model inversion using high-resolution hy-

perspectral airborne imagery–retrieving maize lai to access biomass and grain yield. Field Crops Research 282, 108449.

Kleijn D, Kohler F, Báldi A, Batáry P, Concepción E.D, Clough Y, Díaz M, Gabriel D, Holzschuh A, Knop E, Kovács A, Marshall E.J.P, Tscharntke T, Verhulst J, 2009. On the relationship between farmland biodiversity and land-use intensity in Europe. Proceedings of the Royal Society B: Biological Sciences 276, 903–909.

Klein, N., Theux, C., Arlettaz, R., Jacot, A., Pradervand, J.N., 2020. Modeling the effects of grassland management intensity on biodiversity. Ecology and evolution 10, 13518–13529.

Kolecka, N., Ginzler, C., Pazur, R., Price, B., Verburg, P.H., 2018. Regional scale mapping of grassland mowing frequency with Sentinel-2 time series. Remote Sensing 10, 1221.

Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P.H., Estel, S., Haberl, H., Hostert, P., Jepsen, M.R., Kastner, T., et al., 2013. Challenges and opportunities in mapping land use intensity globally. Current opinion in environmental sustainability 5, 484–493.

Kuhn, T., Domokos, P., Kiss, R., Ruprecht, E., 2021. Grassland management and land use history shape species composition and diversity in transylvanian semi-natural grasslands. Applied Vegetation Science 24, e12585.

Lange, M., Feilhauer, H., Kühn, I., Doktor, D., 2022. Mapping land-use intensity of grasslands in germany with machine learning and sentinel-2 time series. Remote Sensing of Environment 277, 112888.

Lee, J., Jurkevich, L., Dewaele, P., Wambacq, P., Oosterlinck, A., 1994. Speckle filtering of synthetic aperture radar images: A review. Remote Sensing Reviews 8.

Lemaire, G., Gastal, F., 2016. Improved estimation of nitrogen uptake in grasslands using the nitrogen dilution curve (reyes et al. 2015), 35: 1561–1570. Agronomy for Sustainable Development 36, 1–2.

Linsley, R.K., Kohler, M.A., Paulhus, J.L., 1949. Applied hydrology. Technical Report.

Lobert, F., Holtgrave, A.K., Schwieder, M., Pause, M., Vogt, J., Gocht, A., Erasmi, S., 2021. Mowing event detection in permanent grasslands: Systematic evaluation of input features from Sentinel-1, Sentinel-2, and Landsat 8 time series. Remote Sensing of Environment 267, 112751.

Louarn, G., Chabbi, A., Gastal, F., 2020. Nitrogen concentration in the upper leaves of the canopy is a reliable indicator of plant n nutrition in both pure and mixed grassland swards. Grass and Forage Science 75, 127–133.

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L., Merchant, J.W., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data 21, 1303–1330. doi:10.1080/014311600210191.

Luyssaert, S., Jammet, M., Stoy, P.C., Estel, S., Pongratz, J., Ceschia, E., Churkina, G., Don, A., Erb, K., Ferlicoq, M., et al., 2014. Land management and land-cover change have impacts of similar magnitude on surface temperature. Nature Climate Change 4, 389–393.

Ma, Q., Chai, L., Hou, F., Chang, S., Ma, Y., Tsunekawa, A., Cheng, Y., 2019. Quantifying grazing intensity using remote sensing in alpine meadows on qinghai-tibetan plateau. Sustainability 11, 417.

Mahajan, G., Pandey, R., Sahoo, R., Gupta, V., Datta, S., Kumar, D., 2017. Monitoring nitrogen, phosphorus and sulphur in hybrid rice (oryza sativa l.) using hyperspectral remote sensing. Precision agriculture 18, 736–761.

Mahmood, R., Pielke Sr, R.A., Hubbard, K.G., Niyogi, D., Dirmeyer, P.A., McAlpine, C., Carleton, A.M., Hale, R., Gameda, S., Beltrán-Przekurat, A., et al., 2014. Land cover changes and their biogeophysical effects on climate. International journal of climatology 34, 929–953.

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. Biochimica et Biophysica Acta (BBA)-Protein Structure 405, 442–451.

Meyfroidt, P., Chowdhury, R.R., de Bremond, A., Ellis, E.C., Erb, K.H., Filatova, T., Garrett, R., Grove, J.M., Heinimann, A., Kuemmerle, T., et al., 2018. Middle-range theories of land system change. Global environmental change 53, 52–67.

Milionis, N., 2016. The Land Parcel Identification System: a useful tool to determine the eligibility of agricultural land – but its management could be further improved. Special Report 25. European Court of Auditors. 12, rue Alcide De Gasperi, 1615 Luxembourg, LUXEMBOURG. doi:10.2865/39118.

Minet, J., Curnel, Y., Gobin, A., Goffart, J.P., Mélard, F., Tychon, B., Wellens, J., Defourny, P., 2017. Crowdsourcing for agricultural ap-

plications: A review of uses and opportunities for a farmsourcing approach. Computers and Electronics in Agriculture 142, 126–138.

Monti-Guarnieri, A., Manzoni, M., Giudici, D., Recchia, A., Tebaldini, S., 2020. Vegetated target decorrelation in SAR and interferometry: models, simulation, and performance evaluation. Remote Sensing 12, 2545.

Morishita, Y., Hanssen, R.F., 2014. Temporal decorrelation in L-, C-, and X-band satellite radar interferometry for pasture on drained peat soils. IEEE Transactions on Geoscience and Remote Sensing 53, 1096–1104.

Mutanga, O., Kumar, L., 2007. Estimating and mapping grass phosphorus concentration in an african savanna using hyperspectral image data. International Journal of Remote Sensing 28, 4897–4911.

O'Mara, F.P., 2012. The role of grasslands in food security and climate change. Annals of botany 110, 1263–1270.

Ottosen, T.B., Lommen, S.T., Skjøth, C.A., 2019. Remote sensing of cropping practice in Northern Italy using time-series from Sentinel-2. Computers and Electronics in Agriculture 157, 232–238.

Owen, P.W., Milionis, N., Papatheodorou, I., Sniter, K., Viegas, H.F., Huth, J., Bortnowschi, R., et al., 2016. The land parcel identification system: A useful tool to determine the eligibility of agricultural land—but its management could be further improved. Special Report 25.

Pärtel, M., Bruun, H.H., Sammul, M., 2005. Biodiversity in temperate European grasslands: origin and conservation., in: Grassland science in Europe. Grassland Science in Europe, pp. 1–14.

Peeters, A., Beaufoy, G., Canals, R., De Vliegher, A., Huyghe, C., Isselstein, J., Jones, J., Kessler, W., Kirilovsky, D., Van Den Pol-Van Dasselaar, A., 2014. Grassland term definitions and classifications adapted to the diversity of european grassland-based systems, in: 25th EGF General Meeting on "EGF at 50: The Future of European Grasslands, pp. 743–750.

Piipponen, J., Jalava, M., de Leeuw, J., Rizayeva, A., Godde, C., Cramer, G., Herrero, M., Kummu, M., 2022. Global trends in grassland carrying capacity and relative stocking density of livestock. Global Change Biology .

Pimstein, A., Karnieli, A., Bansal, S.K., Bonfil, D.J., 2011. Exploring remotely sensed technologies for monitoring wheat potassium and

phosphorus using field spectroscopy. Field Crops Research 121, 125–135.

Pongratz, J., Dolman, H., Don, A., Erb, K.H., Fuchs, R., Herold, M., Jones, C., Kuemmerle, T., Luyssaert, S., Meyfroidt, P., et al., 2018. Models meet data: Challenges and opportunities in implementing land management in earth system models. Global change biology 24, 1470–1487.

Pullanagari, R., Dehghan-Shoar, M., Yule, I.J., Bhatia, N., 2021. Field spectroscopy of canopy nitrogen concentration in temperate grasslands using a convolutional neural network. Remote Sensing of Environment 257, 112353.

Punalekar, S.M., Verhoef, A., Quaife, T., Humphries, D., Bermingham, L., Reynolds, C., 2018. Application of sentinel-2a data for pasture biomass monitoring using a physically based radiative transfer model. Remote Sensing of Environment 218, 207–220.

Quan, X., He, B., Yebra, M., Yin, C., Liao, Z., Zhang, X., Li, X., 2017. A radiative transfer model-based method for the estimation of grassland aboveground biomass. International Journal of Applied Earth Observation and Geoinformation 54, 159–168.

Rabus, B., Wehn, H., Nolan, M., 2010. The importance of soil moisture and soil structure for InSAR phase and backscatter, as determined by fdtd modeling. IEEE transactions on geoscience and remote sensing 48, 2421–2429.

Radoux, J., Bogaert, P., 2020. About the pitfall of erroneous validation data in the estimation of confusion matrices. Remote Sensing 12, 4128.

Radoux, J., Bourdouxhe, A., Coppée, T., De Vroey, M., Dufrêne, M., Defourny, P., 2023. A consistent land cover map time series at 2 m spatial resolution—the lifewatch 2006-2015-2018-2019 dataset for wallonia. Data 8, 13.

Ramankutty, N., Mehrabi, Z., Waha, K., Jarvis, L., Kremen, C., Herrero, M., Rieseberg, L.H., et al., 2018. Trends in global agricultural land use: implications for environmental health and food security. Annual review of plant biology 69, 789–815.

Ramoelo, A., Skidmore, A., Cho, M.A., Mathieu, R., Heitkönig, I., Dudeni-Tlhone, N., Schlerf, M., Prins, H., 2013. Non-linear partial least square regression increases the estimation accuracy of grass nitrogen and phosphorus using in situ hyperspectral and environmen-

tal data. ISPRS journal of photogrammetry and remote sensing 82, 27–40.

Ramoelo, A., Skidmore, A.K., Schlerf, M., Mathieu, R., Heitkönig, I.M., 2011. Water-removed spectra increase the retrieval accuracy when estimating savanna grass nitrogen and phosphorus concentrations. ISPRS journal of photogrammetry and remote sensing 66, 408–417.

Rapinel, S., Mony, C., Lecoq, L., Clement, B., Thomas, A., Hubert-Moy, L., 2019. Evaluation of sentinel-2 time-series for mapping floodplain grassland plant communities. Remote sensing of environment 223, 115–129.

Reeves, M.C., Baggett, L.S., 2014. A remote sensing protocol for identifying rangelands with degraded productive capacity. Ecological Indicators 43, 172–182.

Reinermann, S., Asam, S., Kuenzer, C., 2020. Remote sensing of grassland production and management—a review. Remote Sensing 12, 1949.

Reinermann, S., Gessner, U., Asam, S., Ullmann, T., Schucknecht, A., Kuenzer, C., 2022. Detection of grassland mowing events for germany by combining sentinel-1 and sentinel-2 time series. Remote Sensing 14, 1647.

Rouse Jr, J., Haas, R., Schell, J., Deering, D., 1973. Paper a 20, in: Third Earth Resources Technology Satellite-1 Symposium: The Proceedings of a Symposium Held by Goddard Space Flight Center at Washington, DC on, p. 309.

Rubo, S., Zinkernagel, J., 2022. Exploring hyperspectral reflectance indices for the estimation of water and nitrogen status of spinach. Biosystems Engineering 214, 58–71.

Rudel, T.K., Schneider, L., Uriarte, M., Turner, B.L., DeFries, R., Lawrence, D., Geoghegan, J., Hecht, S., Ickowitz, A., Lambin, E.F., et al., 2009. Agricultural intensification and changes in cultivated areas, 1970–2005. Proceedings of the National Academy of Sciences 106, 20675–20680.

Sano, E.E., Ferreira, L.G., Asner, G.P., Steinke, E.T., 2007. Spatial and temporal probabilities of obtaining cloud-free Landsat images over the brazilian tropical savanna. International Journal of Remote Sensing 28, 2739–2752.

Saralioglu, E., Gungor, O., 2020. Crowdsourcing in remote sensing: A

review of applications and future directions. IEEE Geoscience and Remote Sensing Magazine 8, 89–110.

Savage, J., Woodcock, B.A., Bullock, J.M., Nowakowski, M., Tallowin, J.R., Pywell, R.F., 2021. Management to support multiple ecosystem services from productive grasslands. Sustainability 13, 6263.

Schuster, C., Ali, I., Lohmann, P., Frick, A., Förster, M., Kleinschmit, B., 2011. Towards detecting swath events in TerraSAR-x time series to establish NATURA 2000 grassland habitat swath management as monitoring parameter. Remote Sensing 3, 1308–1322.

Schwieder, M., Buddeberg, M., Kowalski, K., Pfoch, K., Bartsch, J., Bach, H., Pickert, J., Hostert, P., 2020. Estimating grassland parameters from sentinel-2: A model comparison study. PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science 88, 379–390.

Schwieder, M., Wesemeyer, M., Frantz, D., Pfoch, K., Erasmi, S., Pickert, J., Nendel, C., Hostert, P., 2022. Mapping grassland mowing events across germany based on combined sentinel-2 and landsat 8 time series. Remote Sensing of Environment 269, 112795.

Shahan, J.L., Goodwin, B.J., Rundquist, B.C., 2017. Grassland songbird occurrence on remnant prairie patches is primarily determined by landscape characteristics. Landscape Ecology 32, 971–988.

Shang, J., Liu, J., Poncos, V., Geng, X., Qian, B., Chen, Q., Dong, T., Macdonald, D., Martin, T., Kovacs, J., et al., 2020. Detection of crop seeding and harvest through analysis of time-series Sentinel-1 interferometric SAR data. Remote Sensing 12, 1551.

Sibanda, M., Mutanga, O., Rouget, M., 2015. Examining the potential of sentinel-2 msi spectral resolution in quantifying above ground biomass across different fertilizer treatments. ISPRS Journal of Photogrammetry and Remote Sensing 110, 55–65.

Silva, J.P., Toland, J., Jones, W., Eldridge, J., Thorpe, E., O'Hara, E., 2008. LIFE and Europe´s grasslands, restoring a forgotten habitat. https://ec.europa.eu/environment/archives/life/publications/lifepublications/lifefocus/documents/grassland.pdf. Accessed: 2021-03-02.

Ramos da Silva, R., Werth, D., Avissar, R., 2008. Regional impacts of future land-cover changes on the amazon basin wet-season climate. Journal of climate 21, 1153–1170.

Smith, P., 2013. Delivering food security without increasing pressure on land. Global Food Security 2, 18–23.

Sollenberger, L.E., Kohmann, M.M., Dubeux Jr, J.C., Silveira, M.L., 2019. Grassland management affects delivery of regulating and supporting ecosystem services. Crop Science 59, 441–459.

Statbel, 2015. (direction générale statistique - statistics belgium) - service public fédéral economie, p.m.e., classes moyennes et energie. chiffres clés de l'agriculture. https://statbel.fgov.be/sites/default/files/files/documents/landbouw. Accessed: 2021-07-23.

Statbel, 2020. (direction générale statistique - statistics belgium) - service public fédéral economie, p.m.e., classes moyennes et energie. chiffres clés de l'agriculture. https://statbel.fgov.be/sites/default/files/files/documents/landbouw. Accessed: 2021-07-23.

Strozzi, T., Dammert, P.B., Wegmuller, U., Martinez, J.M., Askne, J.I., Beaudoin, A., Hallikainen, N., 2000. Landuse mapping with ERS SAR interferometry. IEEE Transactions on Geoscience and Remote Sensing 38, 766–775.

Stumpf, F., Schneider, M.K., Keller, A., Mayr, A., Rentschler, T., Meuli, R.G., Schaepman, M., Liebisch, F., 2020. Spatial monitoring of grassland management using multi-temporal satellite imagery. Ecological Indicators 113, 106201.

Sutcliffe, L.M., Schraml, A., Eiselt, B., Oppermann, R., 2019. The lucas grassland module pilot–qualitative monitoring of grassland in europe. Palearctic Grasslands 40, 27–31.

Tälle, M., Deák, B., Poschlod, P., Valkó, O., Westerberg, L., Milberg, P., 2018. Similar effects of different mowing frequencies on the conservation value of semi-natural grasslands in Europe. Biodiversity and Conservation 27, 2451–2475.

Tamm, T., Zalite, K., Voormansik, K., Talgre, L., 2016. Relating Sentinel-1 interferometric coherence to mowing events on grasslands. Remote Sensing 8, 802.

Tampuu, T., Praks, J., Uiboupin, R., Kull, A., 2020. Long term interferometric temporal coherence and dinsar phase in northern peatlands. Remote Sensing 12, 1566.

Taravat, A., Wagner, M.P., Oppelt, N., 2019. Automatic grassland cutting status detection in the context of spatiotemporal Sentinel-1 imagery analysis and artificial neural networks. Remote Sensing 11, 711.

Tetteh, G.O., Gocht, A., Schwieder, M., Erasmi, S., Conrad, C., 2020. Unsupervised parameterization for optimal segmentation of agricultural parcels from satellite images in different agricultural landscapes. Remote Sensing 12, 3096.

Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proceedings of the national academy of sciences 108, 20260–20264.

Tonn, B., Bausson, C., Ten Berge, H., Buchmann, N., Bufe, C., Eggers, S., Fernández-Rebollo, P., Forster-Brown, C., Hiron, M., Klaus, V., et al., 2020. A management-based typology for european permanent grasslands, in: 28TH GENERAL MEETING OF EUROPEAN GRASSLAND FEDERATION, P. Virkajärvi K. Hakala M. Hakojärvi J. Helin I. Herzon V. Jokela S .... pp. 412–414.

Tscharntke, T., Klein, A.M., Kruess, A., Steffan-Dewenter, I., Thies, C., 2005. Landscape perspectives on agricultural intensification and biodiversity–ecosystem service management. Ecology letters 8, 857–874.

Tuck, S.L., Winqvist, C., Mota, F., Ahnström, J., Turnbull, L.A., Bengtsson, J., 2014. Land-use intensity and the effects of organic farming on biodiversity: a hierarchical meta-analysis. Journal of applied ecology 51, 746–755.

Ulaby, F.T., Bradley, G.A., Dobson, M.C., 1979. Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part ii-vegetation-covered soil. IEEE Transactions on Geoscience Electronics 17, 33–40.

Ullah, S., Si, Y., Schlerf, M., Skidmore, A.K., Shafique, M., Iqbal, I.A., 2012. Estimation of grassland biomass and nitrogen using meris data. International journal of applied earth observation and geoinformation 19, 196–204.

Vályi, K., Rillig, M.C., Hempel, S., 2015. Land-use intensity and host plant identity interactively shape communities of arbuscular mycorrhizal fungi in roots of grassland plants. New Phytologist 205, 1577–1586.

Van Vooren, L., Reubens, B., Broekx, S., Reheul, D., Verheyen, K., 2018. Assessing the impact of grassland management extensification in temperate areas on multiple ecosystem services and biodiversity. Agriculture, Ecosystems & Environment 267, 201–212.

Verburg, P.H., Crossman, N., Ellis, E.C., Heinimann, A., Hostert, P., Mertz, O., Nagendra, H., Sikor, T., Erb, K.H., Golubiewski, N., et al., 2015. Land system science and sustainable development of the earth system: A global land project perspective. Anthropocene 12, 29–41.

Verrelst, J., Rivera, J.P., Moreno, J., Camps-Valls, G., 2013. Gaussian processes uncertainty estimates in experimental sentinel-2 lai and leaf chlorophyll content retrieval. ISPRS journal of photogrammetry and remote sensing 86, 157–167.

Voormansik, K., Jagdhuber, T., Zalite, K., Noorma, M., Hajnsek, I., 2016. Observations of cutting practices in agricultural grasslands using polarimetric SAR. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9, 1382–1396. doi:10.1109/JSTARS.2015.2503773.

Voormansik, K., Zalite, K., Sünter, I., Tamm, T., Koppel, K., Verro, T., Brauns, A., Jakovels, D., Praks, J., 2020. Separability of mowing and ploughing events on short temporal baseline Sentinel-1 coherence time series. Remote Sensing 12, 3784.

Waldner, F., Diakogiannis, F.I., 2020. Deep learning on edge: extracting field boundaries from satellite images with a convolutional neural network. Remote Sensing of Environment 245, 111741.

Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R.B., Chang, Q., 2019. Estimating leaf area index and aboveground biomass of grazing pastures using sentinel-1, sentinel-2 and landsat images. ISPRS Journal of Photogrammetry and Remote Sensing 154, 189–201.

Weiss, M., Baret, F., 1999. Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data. Remote sensing of environment 70, 293–306.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. Remote Sensing of Environment 236, 111402.

White, R.P., Murray, S., Rohweder, M., Prince, S., Thompson, K., et al., 2000. Grassland ecosystems. World Resources Institute Washington, DC.

Wilson, A.M., Jetz, W., 2016. Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. PLoS biology 14, e1002415.

Xu, D., Chen, B., Yan, R., Yan, Y., Sun, X., Xu, L., Xin, X., 2019. Quantitative monitoring of grazing intensity in the temperate meadow steppe based on remote sensing data. International Journal of Remote Sensing 40, 2227–2242.

Zalite, K., Antropov, O., Praks, J., Voormansik, K., Noorma, M., 2016. Monitoring of agricultural grasslands with time series of X-band repeat-pass interferometric SAR. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9, 3687–3697.

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., et al., 2021. Esa worldcover 10 m 2020 v100, in: Zenodo.

Zeller, U., Starik, N., Göttert, T., 2017. Biodiversity, land use and ecosystem services—an organismic and comparative approach to different geographical regions. Global Ecology and Conservation 10, 114–125.

Zhang, B., Zhang, L., Xie, D., Yin, X., Liu, C., Liu, G., 2015. Application of synthetic ndvi time series blended from landsat and modis data for grassland biomass estimation. Remote Sensing 8, 10.

Zhao, B., Dai, Q., Han, D., Dai, H., Mao, J., Zhuo, L., Rong, G., 2019. Estimation of soil moisture using modified antecedent precipitation index with application in landslide predictions. Landslides 16, 2381–2393.

Zheng, J., Li, F., Du, X., 2018. Using red edge position shift to monitor grassland grazing intensity in inner mongolia. Journal of the Indian Society of Remote Sensing 46, 81–88. doi:10.1007/s12524-017-0667-9.

Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in landsat imagery. Remote sensing of environment 118, 83–94.