biodiversa+
European Biodiversity Partnership

EUROPEAN PARTNERSHIP

Remote sensing for habitat monitoring

Habitat Pilot report

Co-funded by
the European Union

Document Information

Grant Agreement number	101052342
Project acronym	Biodiversa+
Project full name	The European Biodiversity Partnership
Biodiversa+ duration	7 years
Biodiversa+ start date:	1 October 2021
For more information about Biodiversa+	Website: https://www.biodiversa.eu/ Email: contact@biodiversa.eu LinkedIn: Biodiversa+

Deliverable title	Remote sensing for habitat monitoring: Pilot report
Dissemination level	Public
Authors	Albin Bjärhall (BOZEN), Cuong Ngo (SGAV), Dan Leština (NCA CZ), Hans Gardfjell (SEPA), Jakub Rataj (NCA CZ), Jesper Erenskjold Moeslund (SGAV), Jozef Šibík (SAS), Maria Šibíkova (SAS), Martin Kelko (SAS), Marek Súľovský (SAS), Mona Naeslund (SEPA), Patrik Oosterlynck (VL-O), Pekka Hurskainen (SYKE), Risto K. Heikkinen (MoE_FI), Sara Wiman (SEPA), Sebastiaan Verbesselt (VL-O), Stien Heremans (VL-O), Tamara Kirin (MESD), Toon Spanhove (VL-O), Tytti Jussila (MoE_FI), Vít Ježek (NCA CZ)
Contributors	Domhnall Finch (NPWS), Maria Miladinova (ExEA), Albert Ferré Codina (DTER)
Work package title	WP2 Promote and support transnational biodiversity monitoring
Task or sub-task title	Sub-task 2.6.7: Pilot Habitat
Lead partner	MoE_FI & SEPA
Date of publication	December 2025
Disclaimer	Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. This deliverable may not have been approved yet by the European Commission and may be subject to change.

What is Biodiversa+

The European Biodiversity Partnership, Biodiversa+, supports excellent research on biodiversity with an impact for policy and society. Connecting science, policy and practice for transformative change, Biodiversa+ is part of the European Biodiversity Strategy for 2030 that aims to put Europe's biodiversity on a path to recovery by 2030. Co-funded by the European Commission, Biodiversa+ gathers partners from research funding, programming and environmental policy actors in European and associated countries to work on 5 main objectives:

1. Plan and support research and innovation on biodiversity through a shared strategy, annual joint calls for research projects and capacity building activities
2. Set up a network of harmonised schemes to improve monitoring of biodiversity and ecosystem services across Europe
3. Contribute to high-end knowledge for deploying Nature-based Solutions and valuation of biodiversity in the private sector
4. Ensure efficient science-based support for policy-making and implementation in Europe
5. Strengthen the relevance and impact of pan-European research on biodiversity in a global context.

More information at: https://www.biodiversa.eu/

Table of contents

Table of acronyms

BEC	Biodiversity and Environmental Consultancy
BOZEN	Autonomous Province of Bolzano (Italy)
CAL/VAL	Calibration / Validation
CBD	Convention on Biological Diversity
Copernicus	EU Earth Observation Programme
DEM	Digital Elevation Model
DTER	Department of Territory, Housing and Ecological Transition, Catalonia (Spain)
EBCC	European Bird Census Council
EEA	European Environment Agency
ELY	Economic Development Centre (Elinkeino, liikenne ja ympäristökeskus)
EO	Earth Observation
ESA	European Space Agency
EUGW	European Grassland Watch (inferred project name)
ExEA	Executive Environment Agency (Bulgaria)
GBF	Kunming–Montreal Global Biodiversity Framework
GBIF	Global Biodiversity Information Facility
GBiOS	Global Biodiversity Observing System
GDAL	Geospatial Data Abstraction Library
GEO BON	Group on Earth Observations – Biodiversity Observation Network
GIS	Geographical Information Systems
GSWE	Global Surface Water Explorer
HR	High Resolution
HRVPP	High Resolution Vegetation Phenology and Productivity
IQR	Interquartile Range
L2A	Level-2A Sentinel-2 Product
LiDAR	Light Detection and Ranging
LPIS	Land Parcel Identification System
LUCAS	Land Use/Cover Area Frame Survey
MAIA	Models and Artificial Intelligence for Earth Observation
MCR	Model Class Reliance
MEPGT	Croatian Directorate for Nature Protection
MESD	Ministry of Environment and Spatial Development (Croatia)
MISR	Multi Image Super Resolution
MMU	Minimum Mapping Unit
MNDWI12	Modified Normalized Difference Water Index (SWIR band 12)

MoE_DK	Ministry of Environment of Denmark
MoE_FI	Ministry of the Environment of Finland
MSFD	Marine Strategy Framework Directive
NaturaSat	Habitat mapping and segmentation software
NCA CZ	Nature Conservation Agency of the Czech Republic
NDVI	Normalized Differential Vegetation Index
NDWI	Normalized Difference Water Index
NPWS	National Parks and Wildlife Service (Ireland)
RF	Random Forest
RFE	Recursive Feature Elimination algorithm
RS	Remote Sensing
S2DR3	Sentinel-2 Deep Resolution v3 model
SAGA	System for Automated Geoscientific Analyses
SAKTI	Finnish Spatial database of protected areas (Suojelualueiden kuviotietojärjestelmä)
SAS	Slovak Academy of Sciences
Satlas AI	Deep-learning super-resolution model
SCL	Scene Classification Layer
SEPA	Swedish Environmental Protection Agency
SISR	Single Image Super Resolution
SR	Super Resolution
STR	Soil Temperature Regime (Soil Moisture Index)
SYKE	Finnish Environment Institute
VH	Vertical-Horizontal polarization (direction of travel of an electromagnetic wave)
VL-O	Flemish Government, Environment & Spatial Development
VV	Vertical-Vertical polarization
WFD	Water Framework Directive
WiW model	Water in Wetlands model
WSI	Wildlife Surveys Ireland

Executive Summary

Monitoring of grassland and wetland habitats across Europe relies currently on national field-based methods, which differ in scope, frequency and indicators. The Habitat Pilot explored how remote sensing (RS) methods could strengthen and complement existing EU and national monitoring frameworks by providing scalable, repeatable and habitat-relevant information. Through coordinated testing across several countries, the pilot demonstrated that RS-based approaches can effectively support the assessment of key ecological attributes such as inundation dynamics, vegetation structure, seasonal productivity patterns and the distribution of potentially valuable grasslands. These findings suggest that a European-wide upscaling of selected RS workflows is technically feasible and can contribute to the implementation of the EU Biodiversity Strategy, the Nature Restoration Law and the Global Biodiversity Framework.

The work was implemented between 2023 and 2025 and organised into seven thematic subtasks reflecting key components of an integrated monitoring system: inundation dynamics, grassland detection and typology validation, habitat condition indicators from time series, mapping of open grasslands, vegetation height estimation using drone LiDAR, super-resolution of satellite imagery, and habitat segmentation and classification using NaturaSat algorithm. Each subtask applied iteratively defined methodologies, in most cases executed in multiple countries, supported by shared data structures, regular coordination meetings and three in-person workshops.

The results demonstrate that several RS-based methods show potential for complementing national monitoring programmes, yet also methodological challenges were detected. Simple optical inundation models proved useful across diverse wetland types, but discrepancies between wetness indices and mixed pixels need attention; time-series indicators captured hydrological variability and seasonal dynamics, though standardisation is needed; semi-automatic mapping workflows help to identify potential open grasslands for survey targeting; and drone-based LiDAR combined with deep learning algorithms provided promising information for mapping the fine-resolution vegetation structures. The pilot also showed promise for methods that require further development before operational use, including habitat segmentation, and mapping of fine-scale habitat features with super-resolution. The importance of in situ data based validation was shown in several subtasks, e.g. large-scale grassland mapping and inundation dynamics detection.

Across all subtasks, differences in national datasets, habitat definitions and reference data availability influenced performance. Moreover, RS-related technical features, e.g. cloud cover, seasonal gaps, and sensor limitations were shown to affect indicator stability. Despite these challenges, the transnational collaboration provided a valuable testbed for identifying where harmonisation is feasible and where methodological flexibility must be retained.

Overall, the Habitat Pilot demonstrates that a coordinated portfolio of RS-based approaches can significantly strengthen European biodiversity monitoring. The work highlights the need for shared processing standards, consistent validation protocols, improved training data and cloud-based infrastructures to support large-scale deployment. The pilot's findings support the development of a modular monitoring framework aligned with major policy needs under the EU Biodiversity Strategy, the Nature Restoration Law and the Global Biodiversity Framework.

1. Context and relevance of a transnational biodiversity monitoring scheme

1.1 Context and aim

The grasslands and wetlands of Europe, rich in biodiversity and providing vital ecosystem services, are threatened due to intensive land-use, drainage, and climate change. These global drivers cause loss and deterioration of habitats and alteration of the natural processes. Reliable information on the extent and condition of valuable grassland and wetland habitats is therefore essential yet often difficult to obtain consistently across regions.

Traditional field surveys provide high-quality data, but are often limited in spatial coverage. Advances in remote sensing (RS), including satellite, airborne, and drone-based approaches in combination with traditional field surveys, have opened new opportunities to monitor the landscapes and habitats across large areas in a harmonised manner. However, the potential of remote sensing as a tool for biodiversity monitoring is still under progress. Major challenges persist in ensuring data comparability, methodological harmonisation, and the translation of RS-derived products and indicators into ecologically meaningful metrics for policy use.

To address these gaps, the Biodiversa+ Habitat Pilot was launched under the European Biodiversity Partnership. The pilot focused on grassland and wetland habitats as representative, dynamic (and threatened) systems where the benefits and challenges of RS-based monitoring could be effectively demonstrated. Its overarching aim was to develop and test harmonised methods for mapping and monitoring the condition of these habitats, by integrating RS data with in situ information.

More specifically, the Habitat pilot:

- reviewed and compared existing national and regional approaches to habitat mapping and monitoring,
- tested promising RS-based methods in multiple European regions and different grasslands and wetlands, and
- provided recommendations for harmonised protocols that can inform a transnational biodiversity monitoring framework.

Through these efforts, the Habitat Pilot aimed to contribute to the broader Biodiversa+ ambition of developing more coherent approaches to biodiversity monitoring in Europe. For this aim, we focused on testing how remote sensing methods could complement field-based monitoring schemes and support future harmonisation efforts.

1.2 Link to public policies and monitoring frameworks

Ecosystems and habitats are becoming increasingly degraded worldwide, including Europe. The CBD Global Biodiversity Framework (GBF)[1] and the EU Biodiversity Strategy (EU BDS)[2] aim to reverse this trend through restoration actions, including the EU Nature Restoration Law. Effective

[1] CBD, global biodiversity framework: https://www.cbd.int/gbf
[2] EU biodiversity strategy: https://environment.ec.europa.eu/strategy/biodiversity-strategy-2030_en

planning, management, and assessment of these actions require comprehensive ecosystem and habitat mapping and monitoring.

Remote sensing methods and products are identified as valuable complements for mapping and monitoring biodiversity including obligations under EU policies[3], where the EU Nature Restoration Law particularly emphasises the use of these methods. The Habitat Pilot supports monitoring and reporting obligations under GBF and EU BDS, mainly for terrestrial habitats, particularly wetlands and grasslands. By assessing how remote sensing can complement existing field-based mapping and monitoring, it provides guidance for mapping and monitoring in the following areas under EU legislation:

- The EU Habitats Directives[4], which require Member States to map, monitor and report on the conservation status of listed habitats and listed species and their habitats, including within Natura 2000 protected areas.
- The EU Nature Restoration Law[5], that requires both national plans, monitoring and reporting of restoration efforts, and state of habitats and agricultural landscapes.
- The *EU Soil monitoring law*[6] by complementary information on ecosystem and habitat distribution and condition

The recommendations from the pilot can also be applied for obligations on monitoring and reporting of indicators under the *GBF*[7]. This relates especially to Goal A where the integrity, connectivity and resilience of all ecosystems are maintained, enhanced, or restored, substantially increasing the area of natural ecosystems by 2050, and the indicators needed to address this goal.

In addition, the Habitat Pilot supports emerging European and global monitoring frameworks, such as the *Group on Earth Observations Biodiversity Observation Network (GEO BON)*[8] and its European network *EUROPABON*[9], that proposes a *Global Biodiversity Observing System (GBiOS)* for monitoring biodiversity, with so-called *Essential Biodiversity Variables (EBVs)*. The aim is to connect diverse data sources, including remote sensing, field surveys, and modelling into integrated systems that can track biodiversity trends consistently across regions and scales.

1.3 European/transnational dimension

Biodiversity monitoring is by nature a transnational task. Species, habitats, and ecological processes extend across administrative borders, and many environmental pressures—such as land-use

---

3 Camia A., Gliottone I., Dowell M., Gilmore R., Coll M., Skidmore A., Chirici G., Caimi C., Brink A., Robuchon M., Ferrario I., Earth Observation in Support of EU Policies for Biodiversity - A deep-dive assessment of the Knowledge Centre on Earth Observation, Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/185588,JRC1329 https://op.europa.eu/en/publication-detail/-/publication/bb713e80-e316-11ed-a05c-01aa75ed71a1/language-en

4 EU habitats directive: https://environment.ec.europa.eu/topics/nature-and-biodiversity/habitats-directive_en

5 EU nature restoration law: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1991&qid=1722240349976

6 Directive (EU) 2025/2360 of the European Parliament and of the Council of 12 November 2025 on soil monitoring and resilience (Soil Monitoring Law): https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202502360

7 Indicators for the Kunming – Montreal Global Biodiversity Framework: https://www.gbf-indicators.org/

8 GEO BON: https://geobon.org/

9 EUROPABON: https://europabon.org/

change, eutrophication, and climate impacts—affect ecosystems both at regional and continental scales. Yet, most monitoring efforts in Europe remain organised nationally, using different definitions, indicators, and methodologies. This fragmentation makes it difficult to compare data or to produce consistent assessments at the European level.

The Habitat Pilot explored ways to reduce some of these inconsistencies by testing how remote sensing–based approaches could be applied in different European regions and habitat types. The work brought together partners with complementary expertise and datasets, allowing practical comparison of methods and indicators. Rather than developing a fully standardised protocol, the pilot aimed to identify the sectors where developing ecologically-based mapping and monitoring methods is crucially needed, and where methodological harmonisation is possible and where regional flexibility is required.

2. Current biodiversity monitoring landscape

2.1 Overview of existing monitoring schemes

Biodiversity monitoring of habitats including wetlands and grasslands across Europe is grounded in a long tradition of field-based mapping and monitoring, with remote sensing methods incorporated to varying degrees[10]. The *Habitat Pilot internal review* conducted at the start of the pilot (in Module 1) confirmed this heterogeneity. Across the 11 participating countries, more than 40 different mapping and monitoring approaches were identified, ranging from traditional field surveys and manual photo-interpretation to advanced machine-learning classifications using multi-source Earth Observation data. The adoption of RS tools varied widely: some countries (e.g. Denmark, Finland, Sweden) have integrated EO products into national habitat mapping workflows, while others rely almost entirely on field mapping due to concerns about RS accuracy or insufficient technical capacity.

Several EU-level initiatives contribute to the harmonised data layers, such as LUCAS, the Copernicus Land Monitoring Service, EEA indicator frameworks, but these do not yet translate into a shared operational framework for habitat condition assessment. As also noted in the pilot's internal review, despite a growing number of promising RS applications, most national monitoring schemes still treat EO mainly as a supplementary, not primary, data source, and lack mechanisms to merge RS with existing field-based indicators.

The Habitat Pilot was initiated within this context: to explore how remote sensing methods could complement existing national schemes and to identify which approaches might be realistically harmonised across Europe.

[10] Moersberger, H., Valdez, J., Martin, J. G. C., Junker, J., Georgieva, I., Bauer, S., Beja, P., Breeze, T. D., Fernandez, M., Fernández, N., Brutons, L., Jandt, U., Bruelheide, H., Kissling, W. D., Langer, C., Liquete, C., Lumbierres, M., Solheim, A.L., Maes, J., Morán-Ordónez, A., Moreira, F., Peér, G., Santana, J., Shamoun-Baranes, J., Smets, B., Capinha, C., McCallum, I., Pereira, H. M., Bonn, A. 2023. Biodiversity monitoring in Europe: User and policy needs. Conservation Letters, vol 17, issue 5. https://conbio.onlinelibrary.wiley.com/doi/10.1111/conl.13038

2.2 Technological developments and innovation

Rapid developments in Earth Observation are changing the possibilities for habitat mapping and monitoring. The pilot' internal review highlighted several technological advances that are particularly relevant for grassland and wetland habitats.

Multi-temporal satellite data: Dense time series from Sentinel-1 and Sentinel-2 now enable monitoring of key ecological processes, including moisture dynamics, inundation patterns, vegetation productivity, and phenology. Improved cloud masking, atmospheric correction, and temporal compositing support the consistent extraction of indicators across regions.

Machine learning and deep learning: Many partners are developing or testing machine-learning-based habitat mapping approaches, using random forests, gradient boosting or deep convolutional networks. These methods allow integration of multispectral, radar, LiDAR and ancillary datasets. However, the review emphasised that harmonised training data remain a major bottleneck, limiting pan-European applicability.

Segmentation and object-based analysis: EO-driven segmentation methods offer new possibilities for delineating ecologically meaningful patches and refining habitat boundaries. These methods were among the most promising identified in pilot's Module 1 due to their applicability across regions and their ability to integrate field data.

Time-series indicators for habitat condition: Several Member States use spectral indicators (e.g. NDVI, NDWI) and derived temporal metrics as proxies for management intensity, hydrological regime and biomass change. The review noted strong potential in time-series-based condition indicators, but also highlighted challenges with their ecological interpretation, validation, and cross-site comparability.

High-resolution structural data: Airborne LiDAR and UAV-based photogrammetry are increasingly used to measure vegetation height and structure—key attributes of habitat condition not captured by medium-resolution satellite sensors. These data provide crucial calibration and validation for EO-based indicators.

Super-resolution and data fusion: Emerging deep-learning methods (e.g. Satlas, S2DR3) can synthetically enhance the spatial resolution of Sentinel-2, improving boundary delineation and potentially supporting fine-scale monitoring when high-resolution imagery is unavailable. Their use is still exploratory but was considered promising in the review.

Cloud computing and open infrastructures: Platforms such as Google Earth Engine, openEO, Copernicus Dataspace and national cloud environments now enable large-scale processing of satellite time series and facilitate collaboration across countries. These infrastructures were essential in the Habitat Pilot subtasks.

Despite these advances, the review concluded that the *key barriers are now methodological rather than technological*, including:

- lack of harmonised reference data,
- inconsistent habitat definitions and other challenges in habitat typologies,
- challenges in linking spectral or structural metrics to ecological concepts of "condition",
- and the need for agreed protocols for calibration, validation and uncertainty assessment.

The Habitat Pilot directly addressed these issues by jointly testing inundation modelling, time-series indicators, segmentation and EO-based habitat classification, super-resolution, and vegetation structure measurements across several European regions. Also different validation exercises were carried out to assess the accuracy and potential limitations in the on-going grassland and wetland monitoring approaches and initiatives conducted in the EU. The findings from these exercises contribute to a clearer understanding of how remote sensing can be operationalised within a coordinated European habitat monitoring framework.

3. Proof of concept: the pilot study

3.1 Activities and timeline

The Habitat Pilot was implemented between 2023 and 2025 as part of the Biodiversa+ transnational biodiversity monitoring initiative. Following the completion of Module 1, which focused on reviewing existing remote sensing methods and identifying methodological gaps, it became clear that the subsequent proof-of-concept phase required a modular approach tackling several key issues. The work was thus organised into seven subtasks, each addressing a specific technical or thematic component of habitat monitoring using Earth Observation (EO) data.

This subtask structure allowed partners to work in parallel on different aspects of habitat mapping and condition assessment while maintaining coordination through shared objectives, regular meetings, and exchange of data and methodological insights. The subtasks represented complementary pillars of an integrated monitoring approach, ranging from generalised classification and inundation analyses to advanced methods such as super-resolution modelling and drone-based structural measurements.

The subtasks established after Module 1 were:

1. Indicators for Habitat Condition Monitoring
2. Inundation Mapping
3. EU Grassland Watch Validation
4. Mapping of Open Grassland Habitats
5. Measuring Vegetation Height and Cover with Remote Sensing
6. Habitat Segmentation and Classification – NaturaSat
7. Super-Resolution Enhancement of Satellite Imagery

Each subtask produced its own technical report. Together, they evaluate the feasibility, strengths, and limitations of using EO-based approaches for mapping and monitoring the extent and ecological condition of wetlands and grasslands across Europe.

The work followed a multi-step process:

1. Scoping and review (Module 1): Compilation of existing approaches, data sources, and indicators and identification of key challenges and knowledge gaps.

2. Subtask design: Establishment of focused workstreams to address the technical needs identified during the review phase.
3. Method testing: Application of remote sensing approaches in pilot sites across (where possible) several countries, covering a range of ecological settings and monitoring contexts.
4. Synthesis: Comparative evaluation of methods, identification of complementarities between subtasks, and assessment of scalability and applicability for operational monitoring.
5. Reporting and recommendations: Preparation of subtask reports and the integration and summarisation of their results into the present final pilot report.

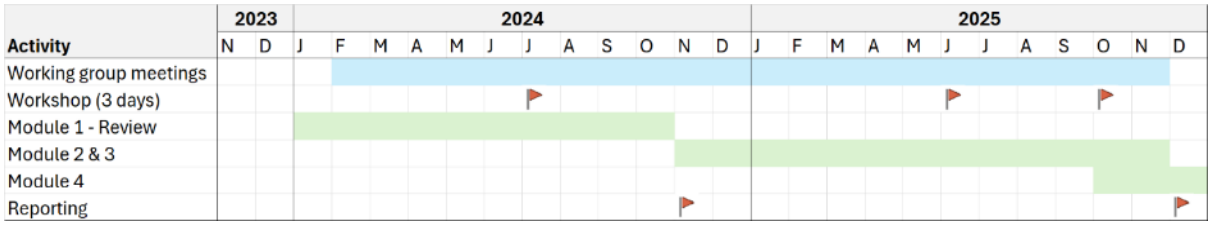
A detailed summary of the methods, results and conclusions from each subtask is presented in Section 3.2, while the full subtask reports are included in the Appendix.

In addition to the work carried out within the individual subtasks, substantial effort was devoted to capacity building and knowledge exchange, both among the pilot partners and through the invitation of external researchers to present projects aligned with the pilot's aims. Partners engaged actively in regular working group meetings and dedicated training sessions designed to facilitate learning and methodological alignment across countries.

Throughout the pilot, working group meetings were held on a weekly or bi-weekly basis, providing a continuous forum for discussion, troubleshooting and coordination. Three in-person workshops were also organised, each attended by approximately 15–20 partners onsite, with additional participants joining online:

- June 2024 – Bolzano, Italy
- June 2025 – Oulu, Finland
- October 2025 – Brussels, Belgium

Table 1 Modules, workshops and reporting timepoints



3.2 Governance and coordination

The pilot was coordinated under the Biodiversa+ partnership, with the Habitat Pilot Coordination Team facilitating communication and integration between national partners. Coordination took place through regular online meetings, dedicated workshops, and joint documentation in a shared workspace.

Each subtask was led by a partner organisation responsible for its thematic or methodological focus. Coordination ensured consistency across subtasks, but partners retained flexibility to adapt approaches according to their national data availability and technical infrastructure.

Decision-making was collegial, based on consensus. The coordination team compiled inputs from all partners into shared deliverables, including the internal review, subtask reports, and this final synthesis.

Data management and analysis were largely decentralised: each partner performed their own processing and contributed summary datasets and figures for synthesis. Cross-validation between subtasks was carried out through joint comparison exercises.

The partners/regions that participated in project as active contributors are:

- BOZEN, Autonomous Province of Bolzano
- MEPGT, Croatia
- MoE_DK, Denmark
- MoE_FI, Finland (project coordination)
- NCA CZ, Czech Republic
- SAS, Slovakia
- SEPA, Sweden (project coordination)
- VL O, Belgium – Flanders
- *Limited participation (Module 1 and 4):*
    - DTER, Autonomous Region of Catalonia (Spain)
    - ExEA, Bulgaria
    - NPWS, Republic of Ireland

3.3 Subtasks – methods, results and conclusions

The structure of the subtasks within the Habitat pilot is illustrated by Figure 1 while TBD shows which partners participated in the different subtasks.

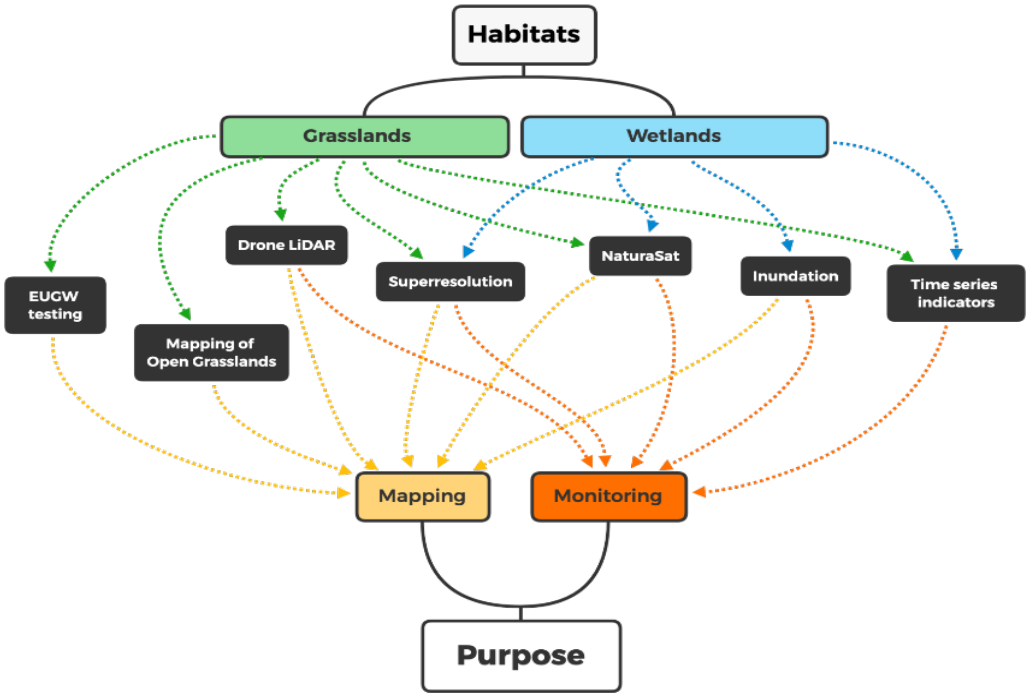


Figure 1 Subtasks within the Habitat pilot. The figure indicates how the subtasks relate to grassland and wetland habitats respectively, and if they focus on mapping or monitoring. Some of them are generic technical methods (time series, super-resolution) which can be used in several applications of habitat mapping or monitoring.

Table 1 Partners involved in the different subtasks. Some of them were limited to a few, while others assembled most of the partners. Only the partners active in Module 2 and 3 are listed here.

Partner country	Inundation	Indicators, monitoring	Super-resolution	EUGW	Veg height	NaturaSat	Grassland mapping
Slovakia	✓	✓		✓		✓	
Finland	✓	✓					✓
Flanders - Belgium	✓	✓	✓	✓	✓	✓	
Bolzano - Italy	✓	✓	✓	✓			
Czech Republic	✓	✓		✓		✓	
Sweden	✓	✓	✓	✓	✓	✓	
Denmark			✓	✓	✓	✓	
Croatia		✓		✓			
No of partners	*6*	*7*	*5*	*7*	*3*	*5*	*1*

3.3.1 Subtask: Indicators for Habitat condition monitoring

Full report in Appendix 1

Aim and rationale

This subtask investigated the potential of multi-temporal satellite data to derive time series of simple, interpretable indicator metrics that are relevant to habitat conditions across multiple habitat types. Focusing on habitat hydrology, our rationale was that wetness and inundation dynamics reflect key ecological processes and changes in them may indicate degradation, recovery, or changing climatic pressures. The work aimed to formulate suitable RS-based hydrological indicators for habitats, test the feasibility across European regions, and clarify how processing choices influence comparability and usefulness for monitoring. In addition to the hydrology cases, the same methodological approach was showcased for monitoring bare-sand cover in dune habitats.

Methods and activities

The workflow was implemented in R and applied to a multi-year collection of Sentinel-2 L2A images retrieved via openEO. The processing steps included:

1. Selection of suitable time windows for indicator calculation, focusing on the growing season.
2. Application of the scene classification layer (SCL) to mask clouds, shadows, snow and other artefacts.
3. Masking out areas of dense tree-cover.
4. Calculation of multiple spectral soil and vegetation moisture indices, including NDMI, NDWI, STR, single-band SWIR and the inundation classifications from other subtasks (e.g. Jussila, WiW).
5. Spatial and temporal gap-filling and smoothing to improve continuity of the time series within a year.
6. Derivation of per-year summary statistics, such as mean, minimum, maximum, and frequency-based inundation metrics.

The workflow was tested in Finland, Flanders, the Czech Republic, Croatia, Bolzano and Sweden, each with different ecological conditions and cloud cover characteristics.

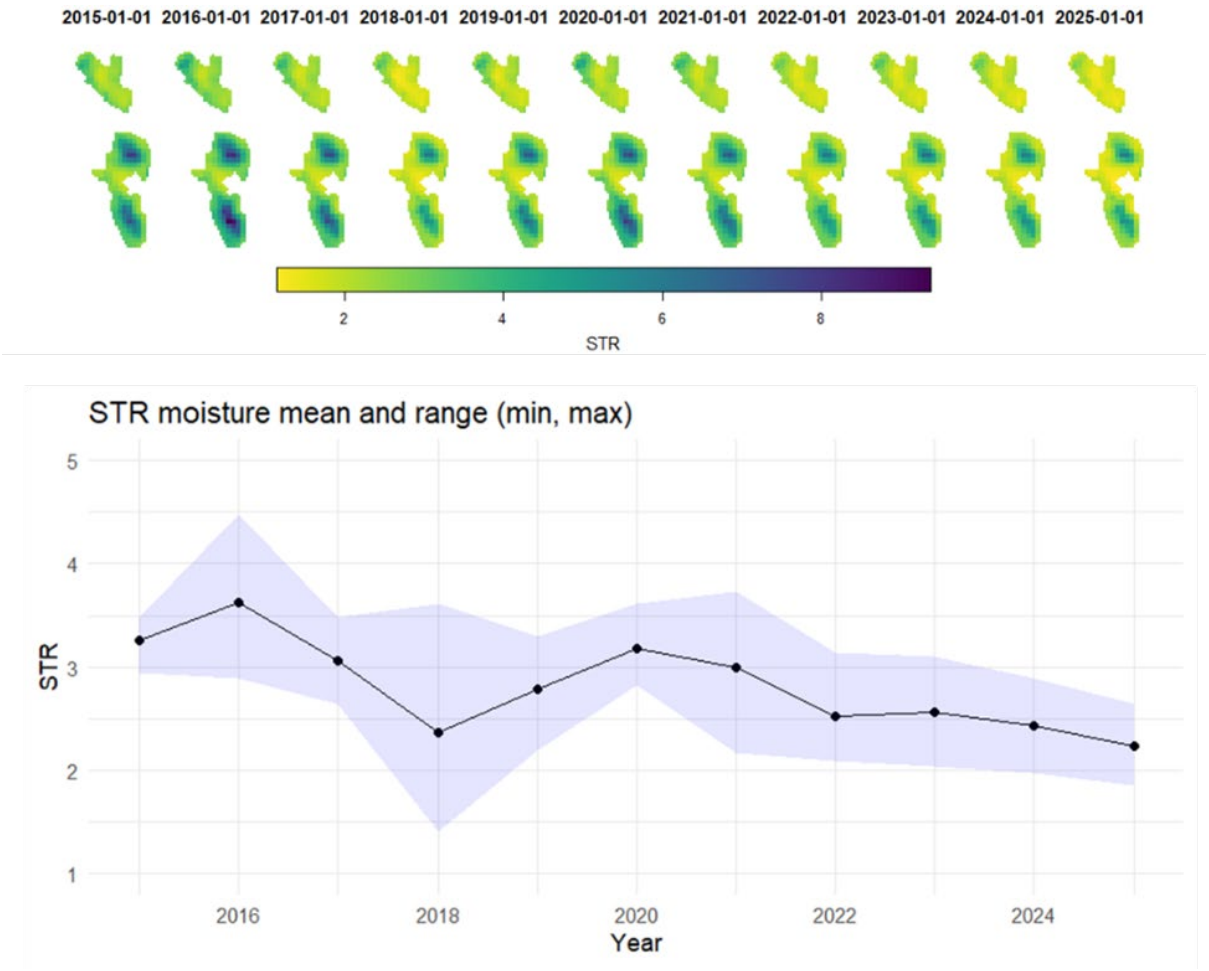


Figure 2. Raster visualisation of soil moisture index STR (5A) averaged per year for each image pixel within the habitat plot, and a graph summary of the plot mean, min and max values per year.

Key results

The time-series analyses revealed considerable variability in data density, with cloud cover strongly affecting observation availability. Many time steps lacked usable images, particularly outside summer months. Temporal completeness varied also depending on the region. Temporal aggregation, smoothing and gapfilling clearly improved the interpretability and comparability of indicators between years and regions. Different choices in preprocessing influenced indicator values, demonstrating that standardised approaches are recommendable to enable robust comparison across regions.

The wetness indices showed divergent and occasionally conflicting seasonal behaviour, as well as varying sensitivity to atmospheric effects, highlighting the substantial influence of index selection on wetness monitoring. Unfortunately, our ability to evaluate the temporal performance of the different indices and inundation indicators was limited by the lack of extensive multi-temporal field validation data.

Conclusions and lessons learned

The subtask demonstrated that Sentinel-2 time series have the potential for deriving harmonised indicators relevant to habitat condition, and that same indicators can serve monitoring across multiple habitat types. However, indicator building requires caution due to time series gaps, atmospheric filtering and effects of preprocessing choices. Seasonal indicators show promise, but their reliability remains uncertain due to the absence of field data for validation. The indicator behaviour also varied between indices and between sites, highlighting the need for harmonised workflows and extensive multi-temporal ground truth measurements to identify which indicators best represent hydrological or ecological conditions in habitats.

Overall, the work provides a valuable foundation for future development of EO-based condition indicators but shows that additional testing, standardisation and validation are required before such indicators can be used operationally across Europe.

3.3.2 Subtask: Inundation mapping

Full report in Appendix 2

Aim and rationale

The aim of this subtask was to explore the potential of optical remote sensing methods to detect and characterise inundation patterns in wetlands and wet grasslands, and to assess how such approaches could contribute to large-scale and potentially Europe-wide inundation monitoring. Inundation—the presence of standing water above the ground surface—is a critical ecological process in many wetland systems. However, its strong spatial variability and rapid temporal dynamics make consistent monitoring through traditional field-based methods inherently difficult. The subtask therefore examined whether spectral information derived from Sentinel-2 images can capture ecologically meaningful wetness gradients and inundation conditions, and how transferable these methods are across diverse European wetland types. The work forms an initial step towards evaluating the feasibility of integrating remote-sensing-based inundation indicators into harmonised habitat monitoring frameworks at continental scale.

Work description

Two decision-tree models were tested: the Water in Wetlands (WiW) model originally developed for the Camargue wetlands in France, and the Jussila model trained in Finnish aapa mires. Both rely primarily on Sentinel-2 near-infrared (band 8) and short-wave infrared bands (bands 11 and 12), with the Jussila model also including the red band and MNDWI12 index. Sentinel-2 L2A images for the study sites were acquired to temporally match the reference data.

Validation sites were situated in seven European countries or regions: Finland, Flanders (Belgium), Slovakia, Czechia, Bolzano (Italy), Croatia and Sweden. These sites included aapa mires, river floodplains and calcareous wetlands. Reference data were generated either through delineation of inundated and dry areas on high-resolution orthophotos or drone data, or through direct field assessment using visual or GPS-tracked wetness observations.

Reference pixels were labelled as inundated, non-inundated or uncertain, and pixel purity was quantified to capture the prevalence of mixed pixels (common in wetlands due to fine-scale

heterogeneity). Uncertain and mixed pixels were excluded from further analyses. We first assessed the feature separability by examining the distribution characteristics of the Sentinel-2 bands and various water-related spectral indices. Then we applied the pre-trained classification models to these new, unseen reference data and evaluated their prediction accuracy.

In Flanders and Finland, we additionally tested the Jussila model on super-resolution images. We measured its effectiveness using standard statistical methods and by comparing the predicted area of standing water to known reference data.

Key results

Across most regions, both inundation models achieved good to very good performance, with macro-F1 scores typically between 0.74 and 0.93. Performance varied by habitat type, reference data quality, and the proportion of mixed pixels.

- In Finland (aapa mires), inundated and non-inundated pixels were well separable using SWIR and NIR bands. Performance was higher with field-based validation than with aerial-image interpretation, reflecting uncertainty in identifying shallow or vegetation-covered inundation from imagery.
- In Flanders, multi-temporal validation demonstrated consistently high performance for both models, although the Jussila and WiW models differ in their sensitivity to water.
- In Czechia and Bolzano, lower performance was linked to limited numbers of inundated pixels or challenging habitats (e.g. reedbeds), highlighting the dependence of accuracy on reference-data representativeness.
- In Slovakia, the models performed well during a large flood event, although small flooded patches were occasionally missed due to insufficient spatial resolution.

Tree-covered pixels and tree shadows were frequent sources of false positives, particularly for the Jussila model, and were masked out before accuracy assessment. Reedbeds and shallow inundation under vegetation remained difficult to detect, reflecting both ecological complexity and limitations of optical sensors.

Super-resolution analysis showed improved visual delineation of small water features, but spectral performance and inundation-area estimates did not improve consistently. In some cases, artefacts introduced by the super-resolved images degraded classification results.

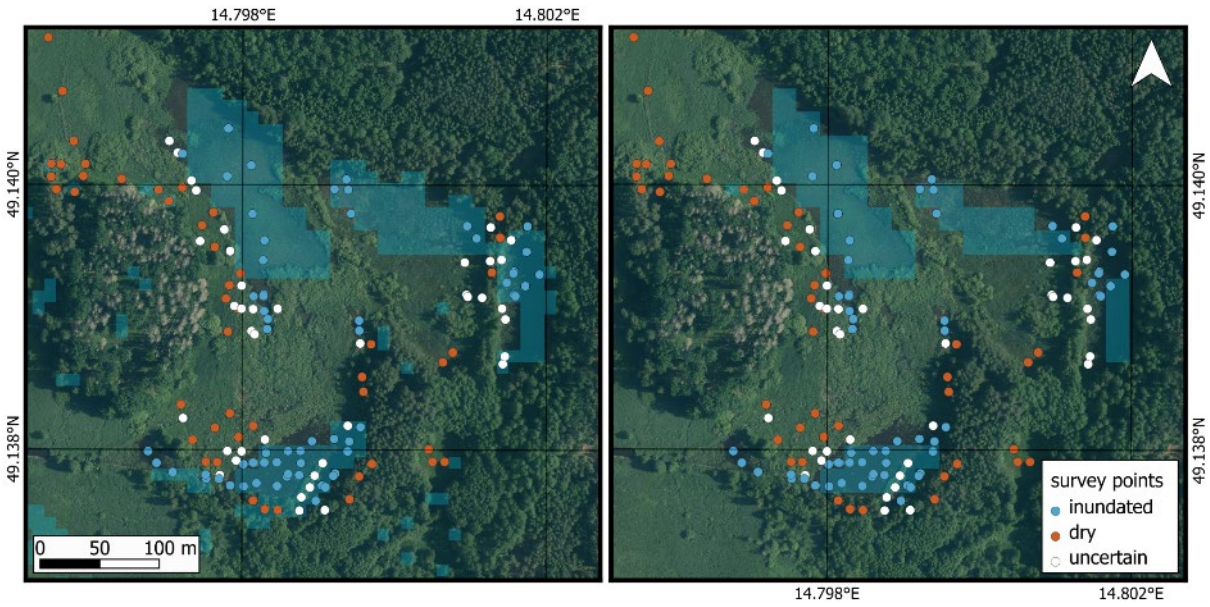


Figure 3. Jussila model output (left) vs. Lefebvre model output (right) for the Kramářka study site in Czechia.

Conclusions and lessons learned

The subtask demonstrated that simple, transferable, Sentinel-2–based inundation models can provide reliable inundation information across diverse European wetland and wet grassland habitats. The models are easy to implement and suitable for large-scale monitoring, but performance depends strongly on the quality and fidelity of reference data. Interpretation uncertainty in aerial imagery, especially in habitats with subtle wetness gradients, remains a major bottleneck for robust validation.

The work also highlighted intrinsic challenges for binary inundation mapping, as many wetland surfaces represent continuous hydrological gradients rather than discrete classes. Mixed pixels, structural vegetation cover and fine-scale patchiness further limit accuracy. Despite these constraints, the models showed promise for capturing ecologically meaningful inundation patterns in open wetland systems.

Future work should expand validation across more regions, refine reference-data collection methods (especially in ambiguous wetness conditions), and integrate complementary satellite data such as Sentinel-1 to address cloud-related limitations. The use of super-resolution can support visual interpretation, but is not yet reliable enough to enhance inundation classification systematically.

3.3.3 Subtask: European Grassland Watch (EUGW) validation

Full report in Appendix 3

Aim and rationale

The aim of this subtask was to obtain an initial understanding of how the current EU Grassland Watch (EUGW) products perform when applied to a set of Natura 2000 sites with contrasting ecological characteristics. The work focused on evaluating both the broad grassland mask and the

EUNIS Level 2 grassland typology, using several complementary validation approaches. The intention was not to assess a final operational product, but to identify strengths, limitations and common sources of error that can inform future development of the EUGW methodology.

Methods and activities

The validation followed four complementary approaches:

1. *Indirect validation using ortho-imagery*
   Partners labelled 400 randomly sampled pixels per site (200 predicted as grassland, 200 as non-grassland) using a harmonised Python-based validation widget, see Figure 4. Decision rules were agreed beforehand to reduce interpretation bias. Challenges included defining sparsely wooded grassland, distinguishing wet grasslands from mires and tall-herb wetlands, and separating permanent grassland from intensively managed grass crops.
2. *Cross-validation with local habitat or vegetation maps*
   For the Belgian SAC “Marais de la Haute Semois”, EUGW grassland classes were compared pixel-by-pixel with a high-quality regional biotope map (Waleunis L1 and L2). Reference classes were reclassified into EUNIS Level 2 categories before comparison.
3. *Direct in-situ validation*
   Field teams in Denmark, Sweden and Czechia validated selected EUGW Typology classes by surveying predicted dry and mesic grasslands in situ. Habitat identity was determined based on plant species composition and structural criteria.
4. *Exploration of temporal trajectories*
   An R-based workflow was developed to analyse grassland-type changes in the EUGW time series (2016–2023), distinguishing persistent from changing pixels and identifying dominant conversion trajectories.

Across all approaches, performance metrics included overall accuracy, precision, recall and F1-score, corrected for the true distribution of grassland versus non-grassland pixels within each site.

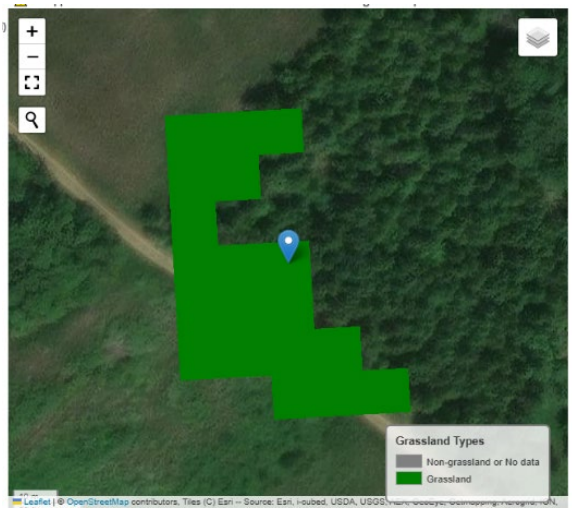


Figure 4. Screenshot of the evaluation widget used for labelling the randomly sampled pixels. The pixel indicated with a marker is the one under evaluation. Model output is provided for the surrounding 49 pixels to allow for contextual interpretation.

Key results

The validation revealed substantial variation in model performance between sites and between grassland types. Broad grassland detection was generally more reliable in landscapes dominated by open vegetation and simpler habitat mosaics. In contrast, performance was considerably lower in ecologically complex areas, such as wooded pastures, wet–dry transition zones, and landscapes with fine-scale variation in moisture or tree cover. These areas also produced the highest numbers of uncertain labels.

Mesic grasslands were the most consistently detected class, although boundaries were often imprecise. Dry grasslands were frequently confused with heaths, sparse rocky vegetation and ruderal communities. Wet and temporarily wet grasslands showed extensive confusion with sedge stands, tall-herb wetland vegetation and mire habitats. The sparsely wooded grassland class (GT7) performed poorly in most regions; predictions were often triggered by scattered trees, forest edges or non-grassland habitats.

Interpretation of ortho-imagery introduced significant uncertainty, especially where vegetation structure obscured the ground layer. Differences between validators were common, particularly for wet grasslands, tree-encroached areas and grasslands on rocky substrates. Mixed pixels and raster edge effects contributed to additional discrepancies.

Exploratory temporal trajectory analysis indicated that many grassland pixels were stable over the observation period, while others showed class transitions. However, given the limitations of the typology component, these temporal patterns should be interpreted with caution.

Table 2 Summary of the accuracy metrics for the evaluated SACs using method 1 for validation of true/false grassland/non-grassland. (Values indicated with an asterisk are uncorrected for grassland/non-grassland proportions)

Country	SAC	#uncertain labels	Overall accuracy	Precision	Recall	F1 score
Sweden	Stora Alvaret	1	0.7560*	0.8416*	0.5838*	0.6894*
Sweden	Tromtö-Almö	36	0.8681*	0.9036*	0.8599*	0.8812*
Sweden	Södra-Stensjö	65	0.7335*	0.8489*	0.7309*	0.7855*
Czechia	Milovice-Mladá	60	0.9678	0.9110	1.000	0.9534
Flanders	Semois	18	0.9428	0.9175	0.9664	0.9413
Flanders	Voeren	7	0.9211	0.8667	0.9713	0.9160
Denmark	Mols Bjerge	12	0.7539	0.9227	0.6273	0.7468
Croatia	Livade uz Bednju II	0	0.9634	0.9500	0.9704	0.9601
Croatia	Stajnicko	1	0.8985	0.9805	0.9912	0.9382
Slovakia	Orava	0	0.9275	0.8693	0.9830	0.9227
Slovakia	Latorica	2	0.9171	0.8200	0.9357	0.8740

Bolzano	Valle di Funes	10	0.9330	0.9684	0.8482	0.9043
Bolzano	Prati dell'Armentara	5	0.9595	0.9798	0.9446	0.9619

Conclusions and lessons learned

The results indicate that the interim EUGW products have potential for broad grassland detection in relatively homogeneous landscapes but may require substantial refinement before they can support operational monitoring of EUNIS Level 2 grassland types. Performance varied widely across ecological settings, and several grassland classes—especially sparsely wooded grasslands and wet grassland types—were not reliably distinguished in their current form. Reference data uncertainties played a major role in several sites, underscoring the need for clearer class definitions and more consistent interpretation guidelines. Improved training data and incorporation of additional environmental variables are likely necessary to enhance typology performance.

Overall, the subtask provides a realistic picture of both the possibilities and current limitations of continental-scale grassland mapping with satellite-based machine learning. It offers a valuable basis for future model development and refinement.

3.3.4 Subtask: Mapping of open grassland habitats

Full report in Appendix 4

Aim and rationale

This subtask aimed to assess whether medium-resolution Earth Observation data, combined with national geospatial datasets, can be used to identify areas with high potential to host species-rich semi-natural grasslands. Because detailed mapping of specific EUNIS classes is currently limited by sparse and heterogeneous field data, the work focused on producing a broader mask of "potential open grasslands" to support survey planning and restoration prioritisation.

Methods and activities

The workflow was developed and tested in Finland, where suitable reference data were available. Sixty mesic semi-natural grassland sites were selected from the national SAKTI database, each embedded within a 2-km analysis buffer. After excluding impervious areas, water and pixels with tree cover above two metres, the remaining open landscape (~26 000 ha) was used for modelling. A supervised Random Forest classifier was trained on point samples representing target grasslands and other open habitats. Predictor variables included multi-season NDVI composites from Sentinel-2, Sentinel-1 backscatter, Copernicus HRVPP parameters, LiDAR-derived canopy and terrain metrics, and elevation. A minimum mapping unit of 0.1 ha was applied to reduce pixel noise, and accuracy was evaluated using independent validation points and comparisons with SAKTI and LPIS datasets.

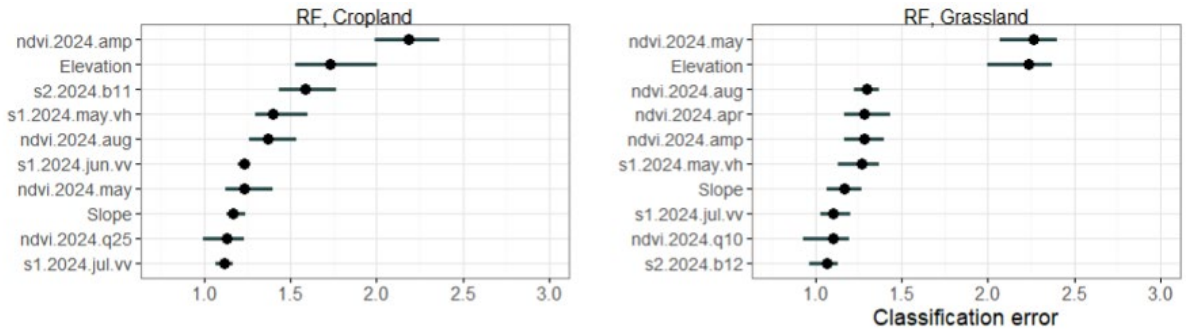


Figure 5. Ten most important features in the RF model for classifying cropland (left) and grassland (right).

Key results

The model achieved an overall F1 score of 0.82, indicating good ability to distinguish potential species-rich grasslands from other open habitats. Agreement with SAKTI mesic grasslands was high: 88% of known grasslands were detected, reflecting the ecological coherence of this class and the relevance of the selected predictor variables. Comparison with LPIS open grasslands showed lower correspondence, largely because LPIS contains a mix of land-use types that are not fully aligned with the target definition.

When evaluated against a combined SAKTI–LPIS mask, the model identified about 70% of known grasslands while also highlighting more than 5 000 hectares of additional areas not captured in existing datasets. Many of these appear plausible as semi-natural grassland candidates and illustrate the added value of EO-based detection for improving the completeness of national inventories. Seasonal NDVI proved particularly important: early-spring and late-summer values differentiated semi-natural grasslands from intensively used fields, while elevation and Sentinel-1 backscatter added contextual and textural information. LiDAR canopy metrics were essential for excluding young forest and shrub cover.

Conclusions and lessons learned

The subtask demonstrates that a semi-automatic EO approach can support national conservation workflows by producing a coarse but ecologically meaningful map of potential open grasslands. The product is not suited for fine-scale boundary mapping or precise estimation of extent, but is effective for screening landscapes, prioritising field surveys and identifying potential restoration targets. Limitations include dependence on representative training data, sensitivity to spectral similarity between habitat types and challenges posed by internal habitat heterogeneity. Transferability to other countries appears feasible where comparable reference datasets and auxiliary geodata exist. Overall, the method provides a practical first step toward integrating EO tools into grassland monitoring and planning.

3.3.5 Subtask: Measuring Vegetation Height and Cover with Remote Sensing

Full report in Appendix 5

Aim and rationale

Vegetation height is a key structural attribute of grassland and wetland habitats, reflecting grazing pressure, shrub encroachment, nutrient status and general habitat condition. Traditionally, height is measured during field surveys, yielding only point-based information. Because the lowest and most ecologically relevant vegetation layers (< 3 m) cannot be reliably detected by satellite sensors, airborne and drone-based LiDAR represents a promising alternative for producing wall-to-wall vegetation height data. This subtask therefore aimed to test a newly developed deep-learning–based method for classifying vegetation height from drone-borne laser scanning data and to evaluate its transferability across countries and ecological contexts.

Methods and activities

Drone LiDAR data were collected at four test sites: Molslaboratoriet and Understed Bakker in Denmark, Örsten in Sweden, and Zammelsbroek in Belgium. Three different drone-mounted LiDAR sensors were used (DJI L1, DJI L2, YellowScan Surveyor Ultra 2), flown during the growing season with consistent flight settings and generating dense point clouds (>1000 points/m²).

All point clouds were processed to colourize points (using DJI Terra or nearest-neighbour colour transfer where misalignment occurred) and then transformed from RGB to CIELAB colour space to better align with human visual perception and model requirements. Data were tiled into smaller, overlapping blocks (40–100 m) to meet computational limits.

The vegetation height model is a 3D semantic segmentation network combining a U-Net architecture with sparse convolutions (SparseConvNet). It uses point coordinates, return information and colour attributes to classify every LiDAR point into four vegetation height classes: bare ground (0 cm), low vegetation (0–25 cm), medium vegetation (25–300 cm) and tall vegetation (>300 cm). The model was trained on data from Molslaboratoriet supplemented with the Hessigheim 3D benchmarking dataset.

For validation, field-measured vegetation heights were collected at GNSS-located points (≥100 per site) and binned into the same height classes as the model output.

Key results

The deep-learning model achieved overall accuracies ranging from 43% to 76% across the four sites. The highest accuracy occurred at Molslaboratoriet (76%) and Zammelsbroek (72%), demonstrating a notable level of transferability despite differences in drone platforms, sensors and processing workflows.

Across all sites, medium vegetation (25–300 cm) was the most reliably predicted class, with F1-scores frequently above 0.7. Low vegetation (<25 cm) proved the most difficult class to classify accurately. Misclassification between low and medium vegetation was common, reflecting both intrinsic spectral ambiguity and probably the challenge of measuring low vegetation precisely in the field. Tall vegetation was expected to be the easiest class, yet accuracy varied widely, indicating that

structural heterogeneity and class boundaries influenced performance. Also, note that the model was trained in a generally open landscape which also might explain this.

The comparison between the Danish training site and the Belgian independent test site was particularly informative: despite major differences in sensors and data-preparation tools, the model produced comparable accuracy, indicating meaningful scalability potential.

Validation uncertainty was significant, partly because field heights had to be manually binned into discrete classes. The report notes that vegetation height is inherently difficult to measure consistently even in the field, and that different binning choices could have produced different accuracy metrics.

Conclusions and lessons learned

The subtask demonstrates that drone-based LiDAR combined with deep-learning segmentation shows clear potential for habitat monitoring, particularly for structural indicators linked to grazing pressure, vegetation encroachment and general habitat condition. The method scaled better than anticipated across countries, suggesting that a harmonised modelling approach may be feasible.

However, several limitations remain. Classification of low vegetation - a class often central to assessing grassland and wetland quality - was unreliable and requires methodological refinement. The approach also depends heavily on field validation, which is itself uncertain for continuous variables like vegetation height. Future efforts should explore the development of validation methods that acknowledges the nature of the parameters vegetation height and cover (varies over the year, cannot be measured consistently in the field), improved class definitions, continuous-height modelling, sensor fusion and expanded training datasets.

Overall, the subtask provides a promising step towards scalable, repeatable, wall-to-wall vegetation height monitoring, while highlighting the need for further development before operational deployment at national or European scale.

3.3.6 Subtask: Habitat Segmentation and Classification - NaturaSat approach

Full report in Appendix 6

Aim and rationale

The aim of this subtask was to test harmonised methods for habitat segmentation and automatic detection of selected Natura 2000 habitat types using the NaturaSat software. NaturaSat integrates multispectral Sentinel-2 data, vegetation plot information and image-processing tools into a single environment designed for habitat exploration and mapping. The objective of the subtask was to evaluate whether the segmentation and classification tools in NaturaSat can be applied consistently across European countries and to assess their performance for detecting habitat boundaries and identifying habitat presence based on relevancy maps.

Methods and activities

Two methodological components were tested: habitat segmentation and pixel-based habitat classification.

For segmentation, the software uses semi-automatic and automatic curve-evolving algorithms applied to Sentinel-2 data. These require an initial location estimate, typically derived from vegetation plots, existing national habitat maps or other point-based habitat observations. Users can adjust parameters related to edge detection, homogeneity thresholds, vector field filtering and stopping criteria. Segmentation results were compared with field-based habitat maps or GPS tracks using mean Hausdorff distances as quantitative metrics.

For classification, the subtask applied the Natural Numerical Network, a deep-learning method that uses statistics derived from all Sentinel-2 optical bands to classify different habitats and compute relevancy maps of their distribution. These maps indicate the likelihood that each pixel belongs to a target habitat, with higher values marking higher relevance. Training data consisted of habitat polygons provided by partners and refined by segmentation where needed. For verification, random point samples from high- and low-relevancy areas were compared with available reference data.

Segmentation and classification were tested using Sentinel-2 data from sites in Flanders (Belgium), Czech Republic, Denmark, Slovakia and Sweden, based on the availability of suitable field observations or recent habitat maps.

Key results

Segmentation:

In the Czech Republic, eleven habitat types were segmented, including alpine heaths (4060), siliceous alpine grasslands (6150), Nardus grasslands (6230), transition mires (7140) and several rocky slope and forest habitats. Mean differences between segmented and field-mapped boundaries were often below 10 metres, which is within or below the Sentinel-2 pixel size. Transition mires (7140) and rocky slope habitats (8110 and 8220) showed the highest agreement, while forests and the 6510 meadow habitat showed larger discrepancies, partly due to their extent and internal variability.

In Flanders, segmentation of 35 mapped water bodies in the Ramsar site Kalmthoutse Heide resulted in a mean Hausdorff distance of 29 metres. Deviations mainly occurred in reed beds, which were included in field mapping but not consistently captured by automatic segmentation.

In Slovakia, segmentation results aligned well with field-derived GPS tracks, with typical mean deviations of one to three pixels.

Classification:

In Denmark, several Natura 2000 habitats were prepared for training, including heaths (4030), coastal dunes (2130), juniper formations (5130) and dry grassland complexes (6210/6230). After excluding classes with insufficient training samples, the Natural Numerical Network produced distinct feature-space separation between selected habitats and generated relevancy maps showing clear habitat-specific gradients.

Validation using random point samples for dry grasslands showed that points in high-relevancy areas were associated with grasslands and low-intensity use categories in the crop farming database in more than 90% cases, whereas low-relevancy points were mainly associated with arable land or non-grassland categories.

Conclusions and lessons learned

The results demonstrate that NaturaSat's segmentation tools can delineate habitat boundaries with good accuracy in all tested regions when suitable point-based occurrence data are available. Mean deviations of one to three Sentinel-2 pixels indicate that segmentation is feasible for many habitat types, although fine-scale or heterogeneous habitats remain challenging. Differences between field-mapped and automatically segmented borders often occurred in ecotones and transitional zones, where even expert delineation can be uncertain. Main advantage of automatic segmentation is the consistency of including or excluding ecotone zones when comparing to the field mapping.

The Natural Numerical Network showed potential for identifying habitat presence and generating relevancy maps, provided when adequate and well-distributed training data are available. The approach was able to capture habitat-specific spectral patterns, but performance varied mainly with training-data quality.

Overall, the subtask indicates that NaturaSat provides a consistent and transferable framework for habitat segmentation and classification across countries, though further optimisation of parameter settings, better harmonisation of habitat definitions and improved training datasets will be necessary before the tool can support operational habitat mapping at European scale.

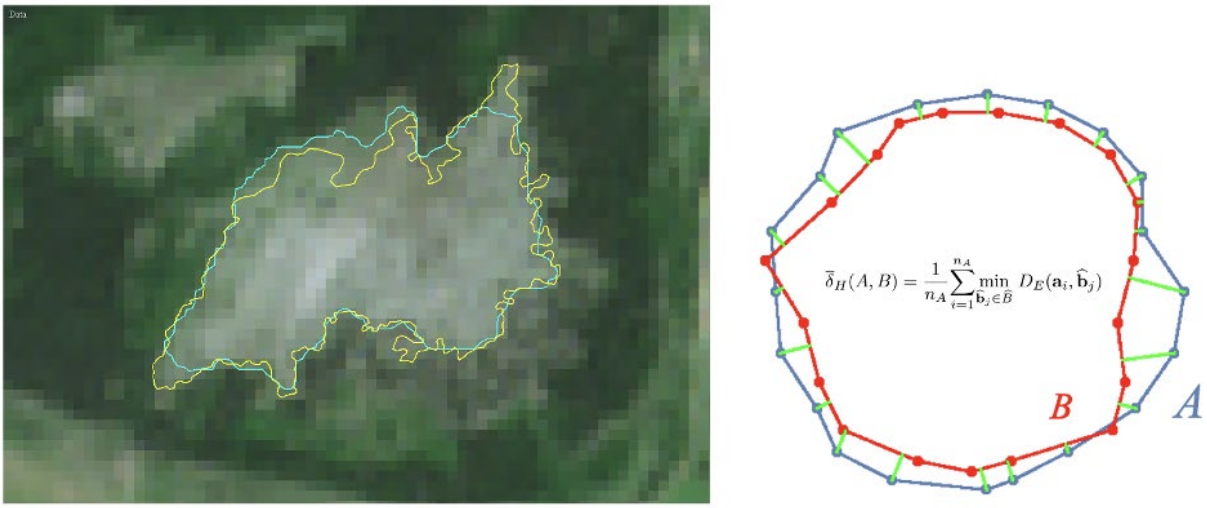


Figure 6. Example of GPS track (yellow) and automatic segmentation (blue) and the formula of Hausdorff distance used for comparison of two curves

3.3.7 Subtask: Super-resolution enhancement of Satellite Imagery

Full report in Appendix 7

Aim and rationale

This subtask explored the potential in using super-resolution (SR) models, designed to enhance the spatial resolution of Sentinel-2 imagery, to support ecological monitoring. The main objective was to assess whether SR outputs can be considered reliable and valuable for improving satellite-based habitat delineation, boundary detection, and reflectance-based analyses in grassland and wetland

ecosystems, where small patch sizes and structural heterogeneity pose challenges to Sentinel-2's native 10–60 m resolution.

Methods and activities

To evaluate the potential of SR products for ecological monitoring, several steps were undertaken. First, an operational pipeline was developed to generate consistent SR mosaics using the pre-trained, open-source multi-image super-resolution (MISR) model Satlas. Next, the performance of Satlas and the single-image super-resolution (SISR) model S2DR3 was assessed both qualitatively and quantitatively by comparing their outputs with Sentinel-2 imagery and local orthophotos across various grassland and wetland sites from multiple pilot partners, as well as and high-resolution satellite data for one site. Finally, S2DR3 outputs were combined with a wetland inundation classification model, and the resulting maps were evaluated against classifications generated from the original Sentinel-2 imagery across the same pilot sites.

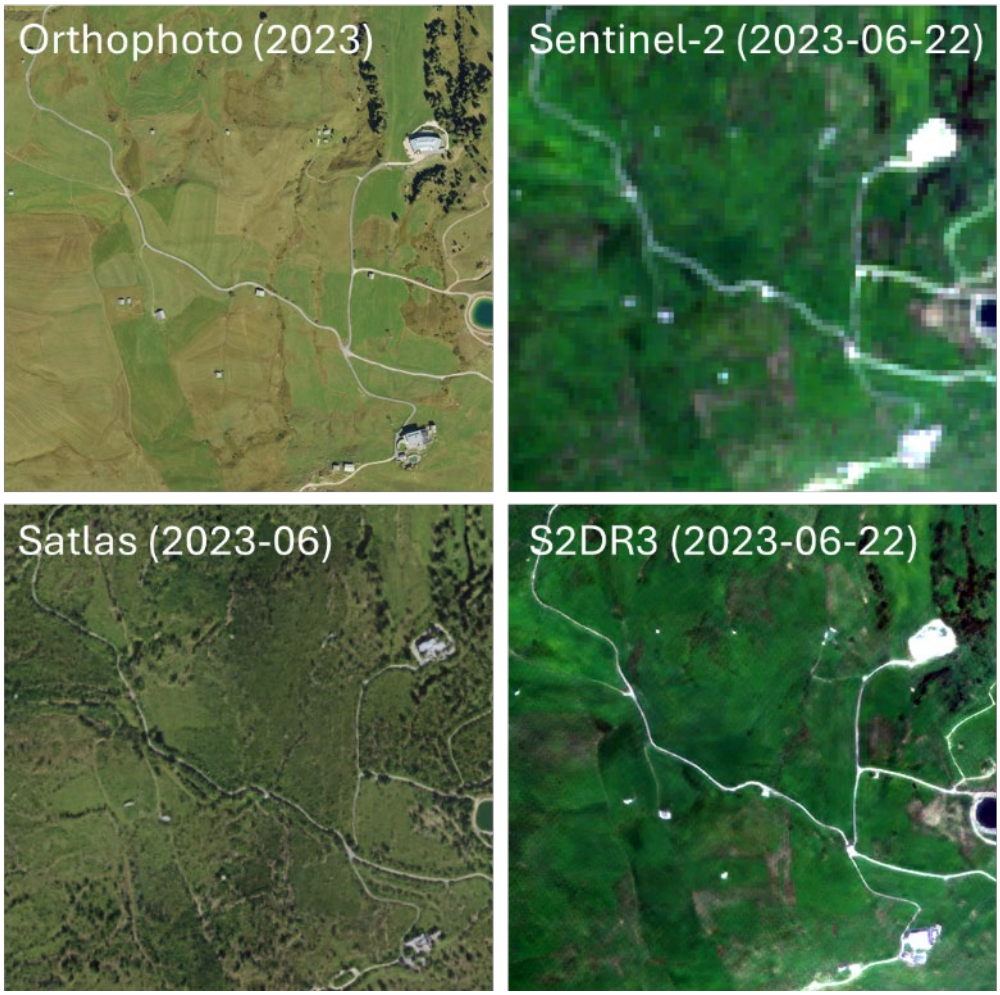


Figure 8. RGB image output from the MISR model Satlas (bottom left, 2.5 x 2.5 m), the SISR model S2DR3 (bottom right, 1 x 1 m) compared to a local orthophoto (top left, 0.2 x 0.2 m) and a Sentinel-2 scene (top right, 10 x 10 m) for an example area in Bolzano.

Key results

Both tested SR models produced visually convincing outputs with enhanced spectral detail compared to the original Sentinel-2 imagery—see figure 4 for a comparison of the RGB output of the two SR models. However, further evaluation using harmonized metrics and standardized benchmarks is essential to establish the reliability of the output.

The SISR model S2DR3, which preserves Sentinel-2's temporal resolution while increasing spatial resolution, demonstrated promising visual results and achieved classification accuracies comparable to those based on original 10m Sentinel-2 data when applied using the inundation model, while offering the added benefit of improved pixel unmixing. Nevertheless, limited availability and lack of model transparency remain key barriers to a potential integration into transnational biodiversity monitoring frameworks.

Conversely, the output of the Satlas model, created using eight Sentinel-2 scenes, exhibited a stronger tendency toward 'hallucinations' and spectral distortions, likely influenced by discrepancies between its training data and the habitats assessed in this pilot.

For both models, quantitative analyses and comparisons with reference datasets were highly variable, reflecting inconsistencies between Sentinel-2 and reference data as well as across datasets themselves. Direct comparisons of the two models are further challenged by the difference in temporal and spatial resolution between the models, and the number of bands included in the model output which will affect not only the calculated metrics, but also the models' usefulness for ecological monitoring.

Conclusions and lessons learned

Super-resolution of Sentinel-2 imagery offers potential for significantly enhancing satellite-based ecological monitoring by providing higher-resolution multispectral data with a high temporal frequency—provided the outputs can be trusted. Our work underscores the challenges of assessing SR model reliability and performance in a robust and comparable way. The findings emphasize the need for high-quality, globally harmonized reference datasets, standardized benchmarks, and comprehensive evaluation metrics that capture spatial accuracy, spectral fidelity, and human perceptual quality of the imagery.

To fully leverage the potential of Sentinel-2 super-resolution, including a potential integration into transnational biodiversity monitoring activities, an openly available SR model trained on high-quality, representative datasets—particularly those reflecting European natural environments—will be essential. Future efforts should therefore prioritize standardized evaluation protocols and stronger alignment with European datasets.

3.4 Budget

The pilot was funded through the Biodiversa+ framework, combining national and European contributions. The overall resources covered personnel time, data access, cloud processing, and coordination activities.

Budget distribution was fixed with equal resources for all partners involved in each Module (1-4) for project work and travelling. Most costs were related to staff time for data processing, analysis,

working group meetings and reporting. Additional expenses for travelling include travel costs for the participation in the three in-situ workshops.

While no major deviations from the planned budget occurred, some differences between partners reflected variations in data processing requirements, methodological approaches, and the degree of engagement in the different project tasks. The pilot's budget was sufficient for exploratory testing but would need substantial scaling to support long-term or continental-level operations.

3.5 Protocols and logistics

The pilot did not follow a single unified protocol, as it was designed to test multiple EO-based approaches. Instead, each subtask applied methods suitable for its specific objectives, using comparable data structures and reporting formats.

Sampling focused on grassland and wetland habitats, selected to represent ecological gradients and data availability across Europe. Sites included both managed and semi-natural systems, covering different hydrological regimes, vegetation structures, and management intensities.

Methods tested included:

- generation of spectral indices (e.g. NDVI, NDWI) time series for vegetation and water dynamics;
- classification of habitat types using multi-temporal imagery;
- analysis of inundation patterns and vegetation seasonality;
- linkage between EO metrics and field-based condition indicators.

Logistical work involved data acquisition from Copernicus Sentinel-1 and Sentinel-2, preprocessing on cloud platforms such as Google Earth Engine or Sentinel Hub, and data harmonisation for cross-comparison between partners.

Although methods differed in detail, the pilot developed a shared understanding of how RS-based indicators could complement national monitoring protocols and provide more consistent information on habitat status.

3.6 Data and sample management

Data handling followed the FAIR principles (Findable, Accessible, Interoperable, and Reusable) as far as possible within the pilot's timeframe. Partners stored and processed EO data locally or on institutional cloud systems, ensuring reproducibility through documented workflows.

Where possible, metadata were standardised to include information on the acquisition period, sensor type, processing steps, and coordinate systems. Summary datasets and key figures were shared through a common repository for synthesis and reporting.

Field data from national monitoring programmes were used mainly for validation and model training. Data sharing agreements varied by country, limiting the extent of harmonisation, but the pilot demonstrated workable models for integrating local and European-level information.

The analyses confirmed that EO indicators such as vegetation productivity (e.g. NDVI), moisture dynamics, and seasonal variability can serve as proxies for several aspects of habitat condition. However, interpretation still depends on complementary ecological knowledge and field verification.

3.7 Difficulties and solutions

Several challenges were encountered during the pilot:

- **Data heterogeneity and comparability:** National datasets differed in structure, classification systems, and accessibility. Regular coordination meetings and the development of a common data template helped mitigate this issue.
- **Technical limitations:** Cloud and snow cover, sensor gaps, and varying spatial resolutions affected data continuity. These were addressed through temporal compositing and testing of alternative data sources.
- **Computational and storage constraints:** EO processing requires significant computing power and storage. The use of cloud-based services and open platforms allowed efficient management of large datasets within limited budgets.
- **Translation to ecological indicators:** Linking spectral metrics to habitat quality requires iterative testing and expert interpretation. Cross-validation between subtasks improved confidence in the results but also showed that further calibration with field data is needed.
- **Coordination across countries:** Different institutional contexts and priorities complicated alignment. Transparent communication, shared documentation, and a pragmatic focus on comparability rather than uniformity proved effective.

Despite these challenges, the pilot successfully demonstrated that a collaborative, multi-source approach is feasible. It built mutual understanding among partners, clarified methodological requirements, and produced a foundation for future development of a harmonised European biodiversity monitoring framework.

4. Outline / Pathway towards harmonised transnational habitat monitoring

This chapter discusses how the findings of the Habitat Pilot can support the development of large-scale European biodiversity monitoring schemes, especially with respect to the ecologically focused, RS-in situ data-based habitat mapping and monitoring. It should be acknowledged that there are multiple potentially useful pathways in this kind of developmental planning process. The Habitat Pilot addressed several relevant methodological questions and approaches for the mapping and monitoring of grassland and wetland habitats, but yet some other, potentially useful methodologies and data issues were left outside the pilot scope due to limited resources. Furthermore, it is imperative to acknowledge that the findings of the pilot can be applied in developing grassland and wetland habitat monitoring towards a harmonised system applicable for many countries, but that these findings can be applied to other habitat types and ecosystems only with caution.

Considering the grassland and wetland habitat RS-in situ data-based monitoring, the outputs of this pilot provide certain key recommendations and general outlines for the future steps in the developmental pathway. In essence, our findings support the importance of addressing several alternative research questions, validation approaches and various mapping and monitoring methods when developing the scheme for the future improved use of RS-in situ monitoring of grasslands and wetlands of conservation concern. Related to this, the pilot results suggest that no single method can sufficiently cover all the different essential features in ecologically aimed grassland or wetland habitat mapping and monitoring. Furthermore, as the understanding and methodological pros and cons of RS-GIS-in situ data-based habitat monitoring approaches still include some insufficiently known features, further studies are inevitably required in building trustful operationalisation of the methods across different countries and various types of grassland and wetland habitats.

To support the development towards harmonised monitoring methods, the Habitat Pilot outlined a key portfolio of complementary approaches, each suited to different ecological conditions, data requirements, and monitoring needs. There are certain main thematic sections in this developmental pathway that need consideration: (i) links and integration of the grassland and wetland mapping and monitoring questions and findings of the pilot with the other corresponding on-going and planned initiatives for the EU level habitat monitoring; (ii) delineating a set of key activities and their potential time phases based on pilot outcomes, to support the future development of large-scale deployment of the tested RS approaches for grassland – wetland monitoring; and (iii) Extraction, comparison and review of the methodologies and results of the on-going other grassland and wetland monitoring projects to verify and complement the results and methodological recommendations of the Habitat pilot.

Overall, the proposal for the general features in the future pathway outlined here emphasises the need for a modular deployment for the modelling approaches, harmonisation across countries, and the gradual expansion of monitoring capacity over time. Yet, at the same time, we acknowledge that the suite of proposed future methodological pathways can be wider than currently estimated.

4.1 General framework of the grassland and wetland habitat monitoring

There are a number of EU level on-going and potential future initiatives and monitoring systems with which the methodological perspectives and findings of the pilot can linked with, and where possible, more intensively integrated to develop useful cross-fertilisation of between the projects. Such initiatives and monitoring frameworks include, for example, the following:

- EU Grassland Watch: The results of the Habitat Pilot have the potential to support the further development of EU Grassland Watch towards improved classification and detection of specifically grasslands of conservation concern, based on the findings on different key RS-data issues for robust identification of ecologically important habitats, and the different ways for including in situ based validation for controlling the quality of the EU Grassland Watch products from the ecological grassland habitats perspective. For this, a more intensive future collaboration between the partners in different grassland habitat monitoring initiatives would be one essential element.

- Potential future EU initiative for wetland mapping: Wetlands are productive and from nature conservation perspective valuable ecosystems that are globally under alarming threats due to land use change and also increasingly due to climate change. Thus, a global initiative, i.e. The Global Wetland Watch, has been launched to provide information to policy practitioners and support the prioritisation and conservation planning of these high-value ecosystems. A corresponding initiative might be useful to be launched in the EU to support the EU-level environmental policies, corresponding to the EU Grassland Watch. For developing such an initiative, the results of the Habitat pilot could provide one useful cornerstone.
- The newly launched World Ecosystem Extent Dynamics (WEED) project: Although WEED has an overall global framework it has via its key partners a central focus on EU, especially as regards developing innovative tools for improved ecosystem mapping and monitoring. The project aims to enhance understanding of ecosystem dynamics and support key environmental policy frameworks, including EU policies. The outputs of the Habitat Pilot can provide useful input for WEED for the methodological and data recommendations for the development of ecologically targeted and robust tools for valuable grassland and wetland ecosystem dynamics.

4.2 Activities and time phases

The activities in making relevant progress towards harmonised large-scale deployment of the grassland and wetland habitat mapping and monitoring approaches – based on the findings of the Habitat Pilot – can be arranged into a set of preparatory actions, and those that are useful in the 'burn-in' phase of the harmonised methodologies implementation. Throughout the developmental pathway there will be often included activities related to administration, protocol development, monitoring, analysis, reporting, and governance.

Key actions in the establishment of common foundations for transnational monitoring include, for example:

- Development of harmonised methodological frameworks for habitat segmentation, hydrological time-series indicators, and for RS-in situ data based habitat identification and quality assessment.
- Establishment of shared validation protocols, building on insights from all the pilot subtasks, complemented by information derived from other on-going related projects.
- Procurement and legal agreements for the use of satellite data, drone acquisitions, and cloud processing environments.
- Development of a centralised technical infrastructure, including standards for preprocessing, gap-filling, metadata, and FAIR-compliant data management.
- Expansion of pilot tests to additional representative areas across Europe.
- Accumulating critical in situ datasets of habitat types occurrences and their key features to be employed as the validation reference data for the RS methods outputs

The ultimate implementation phase of the carefully developed monitoring methods include, e.g., the following features to support scaling up of the monitoring activities across the EU:

- Upscaling of habitat segmentation, identification and quality monitoring methodologies using RS-GIS-in situ data based parameter presets tailored to different habitat types.
- Implementation of annual hydrological indicators using a harmonised openEO-based processing chain, including the integration of radar imagery in cloud-prone regions.
- Operationalisation of the inundation models in a cloud-based environment for monitoring the spatial and ecological changes in water course dynamics use across Member States.
- Capacity building through shared training materials, technical workshops, and support systems.

which then will enable

- Regular production of habitat maps, hydrological indicators, and assessment-ready datasets.
- Direct integration of monitoring outputs into existing EU reporting systems such as Article 17, Article 12, and the Nature Restoration Law.
- Continuous model development through incremental improvements informed by national feedback.
- A permanent coordination structure ensuring consistency, transparency, and long-term continuity.

4.3 Overall linkages of the Habitat Pilot results with existing biodiversity monitoring landscape

The results of the pilot and the potential future work developed based on them may complement, and where possible be integrated with, several existing European initiatives:

- Copernicus programme: The methods in the pilot rely heavily on Sentinel-1 and Sentinel-2, proving a baseline for developing a transnational monitoring programme based on the Copernicus Data Space Ecosystem and harmonised preprocessing pipelines.
- National monitoring programmes: Drone-based LiDAR, field surveys, and local reference layers remain essential for calibration and validation. The results of the pilot can support the development of these national systems.
- Global frameworks: Several indicators studied in the pilot could be aligned with the Essential Biodiversity Variables (EBVs), particularly those related to ecosystem structure and function. These are useful for complementing and exchanging experience on different monitoring approaches with certain global or European initiatives, such as EBOCC, SEEA Ecosystem Accounting and the World Ecosystem Extent Dynamics project (WEED).

EU Grassland Watch: Pilot results show potential but emphasise the need for harmonised validation approaches, consistent interpretation rules, and improved handling of uncertainties. The future monitoring scheme could incorporate EU Grassland Watch as a core component while strengthening its validation framework through the methods developed in this pilot.

4.4 Governance and coordination

A stable governance structure is needed to ensure long-term consistency in biodiversity monitoring across Member States. EuropaBon has proposed an EU Biodiversity Observation Coordination

Centre (EBOCC)[11] for EU coordination of biodiversity monitoring. Biodiversa+ have continued the work on how the hubs as national or sub-national counterparts could be set up, including discussions with JRC (as a member of EuropaBON consortium), EEA and GBIF as bodies that support the compilation of biodiversity monitoring data [12].

Following this work, at the EU level, coordination would fit into the proposed EBOCC or could be implemented in existing policy framework, where the European Environment Agency in collaboration with the Joint Research Centre could be one possibility. This coordination would include responsibility for standard setting, management of shared infrastructure, and overall quality assurance.

At national level, national hubs to EBOCC or existing national governance structures for biodiversity monitoring[13], would provide reference data, conduct field validation, oversee drone operations, and interpret local outcomes.

At regional or local levels, hubs to EBOCC or existing authorities would support field surveys, validation activities, and communication with landowners and stakeholders.

This tiered structure allows consistency across Europe while respecting national responsibilities and local ecological knowledge.

4.5 Protocols, data and sample management

Based on the pilot outputs, the forthcoming developmental work for the harmonised transnational grassland and wetland monitoring programme could consist of a portfolio of interoperable protocols, including e,g, the following: (i) Habitat segmentation protocol such as the one tested within the NaturaSat framework with using habitat-specific presets; (ii) Habitat identification model built upon the baseline procedures of EUGW project, but trimmed towards more effective separation of habitats of conservation concern and in situ validation; (iii) Habitat quality parameter models, e.g. using inundation mapping protocol combining optical and radar observations, built upon tested inundation and hydrological model; (iv) Improved hydrological time-series protocols using harmonised preprocessing, gap filling, and indicator extraction; and (v) Common validation protocols, both for initial model development phase and for their long-term fine-tuning.

Data management should follow FAIR principles and rely on a centralised cloud platform, ensuring consistency and traceability. Sentinel data, inundation results, segmentation outputs, and derived indicators should be stored through shared European infrastructures. Drone and field data should

---

11 Proposal for an EU Biodiversity Observation Coordination Centre (EBOCC): https://op.europa.eu/en/publication-detail/-/publication/bb713e80-e316-11ed-a05c-01aa75ed71a1/language-en https://knowledge4policy.ec.europa.eu/news/proposal-eu-biodiversity-observation-coordination-centre-ebocc_en

12 Vihervaara P. et al. 2023. Strategic document on the options towards governance structure of transnational biodiversity monitoring schemes, national Biodiversity Monitoring Coordination Hubs, European Biodiversity Monitoring Coordination Centre (BMCC) and other relevantinitiatives (PHASE I). https://www.biodiversa.eu/wp-content/uploads/2023/05/D2.8-Biodiversity-monitoring-strategic-Phase-I-report.pdf

13 Petteri Vihervaara, P., Basille, M., Mandon, C., Suni, T., and Lipsanen, A. (2023) Report on the mapping of ministries, agencies and organisations that fund and steer national/ sub-national biodiversity monitoring schemes.

be stored nationally but aligned with common standards. Metadata templates, version control, and transparent processing chains are essential for reproducibility.

4.6 Quality control

Quality assurance should include automated technical checks, model-based performance assessments, expert review, and periodic external audits. This multi-level approach ensures that all stages of monitoring—from field protocols to satellite data processing—meet the standards required for scientific credibility, policy relevance, and long-term comparability.

Appendices

Appendix 1. Indicators for Habitat condition monitoring

Partners involved:

- Jussila Tytti tytti.jussila@syke.fi (Finnish Environment Institute - Suomen ympäristökeskus - Syke)
- Verbesselt Sebastiaan Sebastiaan Verbesselt (Research Institute for Nature and Forest - INBO)
- Stien Heremans Stien Heremans (Research Institute for Nature and Forest - INBO)
- Bjärhall Albin albin.bjaerhall@eurac.edu (Eurac Research)
- Ježek Vít vit.jezek@aopk.gov.cz (Nature Conservation Agency of the Czech Republic - AOPK ČR)
- Gardfjell Hans hans.gardfjell@slu.se (Sveriges Lantbruksuniversitet - SLU)
- Adler Sven sven.adler@slu.se (Sveriges Lantbruksuniversitet - SLU)
- Sibikova Maria maria.sibikova@savba.sk (Slovak Academy of Sciences - SAS)
- Jozef Šibík jozef.sibik@savba.sk (Slovak Academy of Sciences - SAS)
- Tamara Kirin Tamara.Kirin@mzozt.hr (Institute for Environment and Nature, Ministry of Environmental Protection and Green Transition in Croatia)

Background

Successful application of satellite-based information can provide unprecedented possibilities for biodiversity monitoring, offering cost-efficient, harmonised, temporally continuous, and spatially extensive information on variables that can be linked to the ecological condition of habitats. Satellite remote sensing (RS) is especially suitable for monitoring ecosystem level phenomena and features, such as ecosystem function or structure-related variables, as outlined in the Essential Biodiversity Variable (EBV) framework. RS-based habitat condition indicators have the potential to support, alongside with traditionally field-based monitoring, policy and conservation goals under frameworks like the EU Habitats Directive, Ramsar convention or CBD post-2020 targets, EU Strategy for the Danube Region (EUSDR), and Nature-based solutions monitoring, climate adaptation, and carbon accounting.

The recent decades of RS development have provided a plethora of indices and models that correlate with various bio-physical variables on the Earth's surface, such as properties of vegetation, soil or land-cover. Despite a strong theoretical basis, however, the paths from raw spectral observations to ecologically meaningful habitat monitoring products are still not straightforward, and many existing RS-methods are yet to be thoroughly tested in the context of habitat monitoring. Moreover, the practical application of any method for monitoring requires multiple steps from method selection to data (pre)processing and output interpretation. Many of these steps involve methodological challenges and decisions that must be carefully considered. Due to this multitude of possible workflows, efforts for harmonisation are necessary. Also, ideally, these considerations take into account the variability of the information needs in different habitat types and regional realities.

Within this subtask, we piloted an approach for harmonised yearly monitoring of wetland and grassland habitat condition indicators using Sentinel-2 based indicators, with one additional showcase for vegetation-cover changes in dune habitats. For wetlands and grasslands cases, our focus was on hydrological variables (soil moisture and inundation), for which continuous satellite observations are particularly useful due to the temporally dynamic nature of water-regimes in many of the habitat types. Natural hydrology is a core aspect of habitat functioning, particularly in wetlands and mires, and it is currently under pressure from both climate and land-use change. As a key target for ecosystem restoration, monitoring the hydrology is thus of high priority. In this pilot, we first mapped the common needs for hydrological information across habitat types and regions. We then proceeded to test the complete workflow of generating harmonised, yet locally tailored and ecologically meaningful time series indicators for monitoring eco-hydrological conditions.

In dune habitats, the monitoring of bare sand was targeted as an habitat quality indicator. Bare sand is a critical substrate in coastal dune ecosystems. It facilitates aeolian sediment transport that sustains dune morphology and provides open habitat patches for pioneer plants and invertebrates. A decline in bare-sand extent typically signals vegetation encroachment, reduced disturbance, altered sediment supply, or the cessation of traditional management. Restoration commonly re-establishes bare sand by removing vegetation and lightly reworking the surface.

The work on the different case studies (habitat hydrology cases, dunes) demonstrates the potential of optical RS with Sentinel-2 for harmonised monitoring of habitat condition indicators, but also identifies limitations, and highlights areas where further research is needed for the selected cases.

Recommendations

This pilot demonstrates a path from raw satellite observations to harmonised, yearly time series of meaningful habitat condition indicators. From our results, six key recommendations can be drawn for developing operational monitoring systems for habitat conditions from (optical) satellite time series.

1. **Spatial and temporal requirements according to purpose.**

   The monitoring design should be adapted to acknowledge the required spatial and temporal detail. These requirements can only be formulated successfully when there is cooperation between ecologists and remote sensing scientists, each with their domain expertise. Pixel-level products (such as monthly or yearly rasters or change maps) are well suited for detecting local changes and serve as a flexible basis for aggregating information to higher spatial levels. Aggregated site- or region-level indicators make temporal trends easier to interpret by presenting them as time-series plots. Raster- and plot-based products complement each other and are ideally provided together. Temporal aggregation level - such as bi-weekly, monthly, seasonal, or yearly indicators - must be fit-for-purpose. Yearly duration of inundated period, for example, is ideally derived from higher-than-monthly temporal resolution, in order to detect moderate changes in duration. Gaps in time steps, however, increase with temporal detail due to occasional low availability of cloud-free scenes.

2. **Data quality and transparency throughout the workflow.**

   Atmospheric disturbances and data gaps in many regions are unavoidable with passive optical satellites. Reliable monitoring requires transparent pre-processing, reproducible workflows, and documentation of the completeness and uncertainty of the time series. Due to the irregular distribution of observations in time, gap-filling approaches (i.e., temporal aggregation and interpolation, data fusion) should be applied prior to calculation of yearly indicator metrics, to make the yearly summary statistics comparable. The quality of wetness-related observations is also affected by the degree of coverage of the land surface: habitats with tree-cover or dense reedbeds are a challenge for Sentinel-2 remote sensing, and require further methodological development and targeted validation. We recommend masking out tree-covered areas before calculating the wetness time series.

3. **Selection and validation of RS-indicators.**

   The selection of the spectral indicators and indices has a significant effect on the resulting time series. Indicators vary in their sensitivity to chemical, physical and biological conditions of vegetation and soil (like soil moisture), as well as secondary factors such as vegetation composition, structure and seasonal phenology. Extensive ground truth measurements—ideally covering full seasonal phenological/hydrological cycle and drought periods—are essential to evaluate indicator performance and guide the selection of the most suitable indicator(s) for each habitat type. This selection should also reflect the type of information needed for the habitat type: some indices are correlated to moisture-related vegetation health, some are more impacted by soil water content and others are more impacted by bare sand. Indicators can be either solely spectral metrics (indicative of a physical property like soil moisture) or actual modelled physical variables. The first option (piloted in this study) is simple and easy for time series generation, but the latter is easier for interpretation and enables visualisation of indicator uncertainty and confidence level intervals in the end products. However, literature and available models on the

relationships between spectral indices and the complex physical, chemical and biological interactions in natural habitats are scarce and would require further efforts to apply in monitoring. A network of ground-based sensors, long-term monitored vegetation plots, e.g, would complement a RS monitoring programme and improve reliability.

4. Improving harmonisation across regions and habitat types.

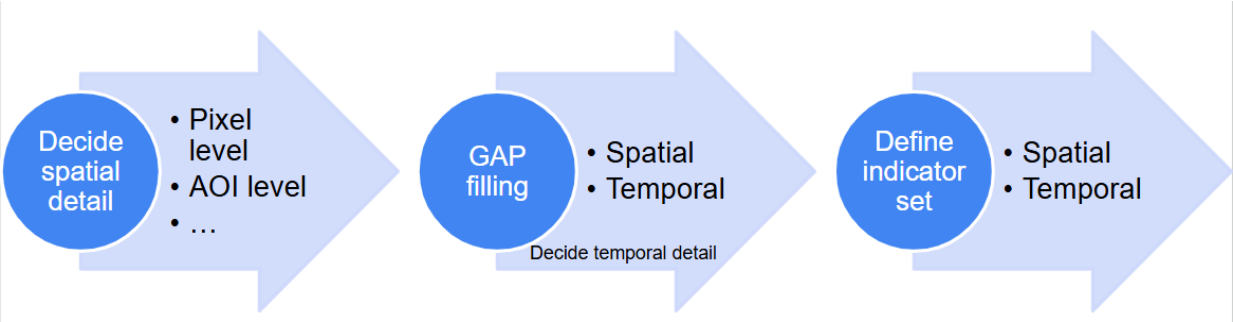
Europe-wide monitoring requires common definitions, processing chains, and validation protocols. Shared reference datasets, open-source pipelines, and cross-country calibration frameworks should be developed under existing initiatives such as Copernicus or EEA's EIONET. Similar RS-based indicators can be applicable and harmonised to be used in multiple habitat types, but indicator interpretation, favourable levels, and level of observation uncertainties are often habitat-specific. Indicator interpretation and application to habitat condition assessment should be harmonised within a habitat type across regions. Regional realities such as relevant seasons for monitoring should be acknowledged.

5. Enhancing scalability through cloud computing and shared infrastructures.

Upscaling from local pilots to national or continental monitoring demands interoperable data infrastructures, harmonised metadata, and efficient access to cloud or high-performance computing. Long-term funding is a requirement for operational monitoring and helps to sustain the infrastructures. Frameworks such as Copernicus Data Space Ecosystem show promise for scalable analysis, but further development in their implementation is needed to optimise visualisation, machine-learning and custom method integration, and cost-efficient large-scale processing. Additionally, to maximize the effectiveness of platforms such as openEO for habitat monitoring, government agencies must be equipped through comprehensive training, capacity building, and increased awareness.

6. Linking indicators to condition assessment frameworks.

Finally, the full value of RS-based indicators will be realised only when they can be plausibly interpreted in terms of habitat condition status. This requires developing habitat-type-specific reference levels or thresholds for favourable moisture or inundation regimes, derived from historical or regional datasets. Integrating RS-based indicators with habitat mapping and ecological interpretation will enable their effective use in assessing condition, restoration success, and policy-relevant reporting.



Work description

Methods and data

Relevancy assessment of indicators for habitat types

To map the requirements for hydrological information in different habitat types, understanding was gathered (i) during discussions throughout the pilot, (ii) the Habitat pilot workshop organised at Oulu in 2025 and (iii) via a spreadsheet-based consultation whereby ecological experts identified the most relevant indicators and metrics linked to inundation and soil/vegetation moisture for grasslands and wetland habitats, together with other RS-relevant information of the habitat types.

The survey had the following sections to be filled by each partner for each of the habitat types occurring in their region:

- Habitat Directive Annex 1 Habitat type
- Relevant season(s) for monitoring
- Required time window for observations to capture ecologically relevant hydrological events in the habitat (quick- versus longer term flooding, prolonged drought periods)
- Expected limitations for satellite-based monitoring (tree-cover, thick vegetation layer, high-resolution heterogeneity, small size of habitats)
- Relevancy of the suggested yearly indicator metrics including the following:
    - Moisture indices:
        - Yearly mean, mean of specific season, wetness minimum, wetness maximum
    - Inundation
        - Number of inundated months or weeks
        - Seasonally/constantly inundated area
        - Ending time of flooding season
    - Drought frequency over multiple years

Sentinel-2 data acquisition

The Copernicus openEO cloud platform was utilised for efficient image acquisition to the study sites, without need to download full image tiles. Multi-year Sentinel-2 L2A image time series were downloaded from openEO with R software. Within the R-script applied in the study, a user can define the area of interest (AOI), the time period (start and end date) and the required Sentinel-2 bands.

The output, image time series datacubes as NetCDF files, was downloaded locally to the user's computer. This method requires the user to have an account on the Copernicus Dataspace Ecosystem (CDSE). Once logged in, no additional API's are required. The time series acquisition

script can be found on GitHub[14]. Using the openEO-based data pipeline developed in this pilot, Sentinel-2 scenes for relevant areas and time-periods can be retrieved automatically.

Building indicator time series

The downloaded image datacubes were processed locally with R into yearly hydrology indicator time series relevant for monitoring through the following steps:

1. Spatial clipping to the area of interest (delimited by, for example, habitat borders, larger biogeographical area, or country).
2. Temporal clipping to the relevant season(s). Varies between regions depending on climatic features such as snow cover conditions and the length of relevant season
3. Filtering observations to exclude low-quality and non-target areas: areas featured by clouds, shadows, and snow (SCL), or non-target land cover, e.g. urban areas and trees.
4. Spatial gap filling (spatial interpolation).
5. Calculating indices and/or machine learning classification (NDMI, NDWI, STR, SWIR, NDPI, water detection classification, MSAVI for dunes)
6. Temporal aggregation for within-year observations (e.g. bi-weekly or monthly).
7. Reporting spatial and temporal completeness of the timeseries based on raw datapoints.
8. Temporal gap filling using temporal interpolation methods.
9. Temporal aggregation to yearly indicators.
10. Visualisation: graphs, raster time series

Several partners adapted this approach to their own study area, habitats and indicators, resulting in variations to the steps listed above. An overview of the cases together with the reference to the scripts and datasets is given in A1_Table 1.

A1_Table 1 An overview of the different study areas, monitored habitats, processing steps, habitat quality indicators and link to the datasets and code. Jussila model = The classification tree model developed for water-covered vs dry surfaces detection in aapa mires in Jussila et al. (2024).

10	Number of test sites + location	Period	Biophysical indicator	Habitats	Processing pipeline	Habitat quality indicator	Accessibility of code, results and dataset
INBO	Flanders (Belgium) Demervallei, 3 test sites (Webbekomsbroek, Schulensmeer, Kloosterbeemden), 10 habitats.	01/01/2020 - 13/10/2025	Inundation, soil moisture	7140 (3x), 6510 (7x)	Used steps: 1, 3, 4 (majority voting with 3x3 kernel), 5 (Wiw, Jussila model, NDMI, NDWI, STR, SWIR1 (B11),	2-weekly or monthly monitoring. Most relevant season: spring - autumn. Min, mean, max	Github page: https://github.com/inbo/BiodiversaHabitatPilot/tree/master

[14] https://github.com/inbo/BiodiversaHabitatPilot/blob/master/Habitat_condition_indicators/source/hydrology/analysis/Retrieve%20Datacube%20OpenEO%20-%20Demervalley.Rmd

					NDPI, 6 (monthly, min, mean, max aggregation), 7, 8 (linear interpolation), 9,10	values of moisture indices. Number of inundated timesteps. % of constantly inundated, seasonally inundated and permanently dry areas.	
NCA CZ	Třeboňsko Protected Landscape Area (Czechia), 2 test sites (Kramářka, Stará a Nová řeka)	01/01/2020 - 13/10/2025	Inundation, soil moisture	7140, 3150	Used steps: 1,2 (clip to April-September), 3, 5 (Wiw, Jussila model, NDMI, NDWI, SWIR1 (B11), 6 (monthly, min, max aggregation), 9,10	2-weekly or monthly monitoring. Most relevant season: spring - autumn. Min, mean, max values of moisture indices. Number of inundated timesteps. % of constantly inundated, seasonally inundated and permanently dry areas	
MoE_FI	n=20 aapa mire sites in Pohjanmaa region in Finland. Full analysis for one site	2015-2025	Inundation, soil moisture	7310 (20x)	Used steps: 1, 3, 4 (majority voting with 3x3 kernel), 5 (Wiw, Jussila model, NDMI, NDWI, STR, SWIR1 (B11), NDPI, 6 (bi-weekly and monthly, min, mean, max aggregation), 8 (linear interpolation), 9, 10	2-weekly or monthly monitoring. Most relevant season:late spring - early autumn. Yearly min, mean, max values of moisture indices. Number of inundated timesteps. % of constantly inundated, seasonally inundated and permanently dry areas.	Github page: https://github.com/tyttijussila/habitatpilot2025/tree/main
MEPGT	Croatia, SAC HR2001305 Zvečevo, molinia grassland	2015-2025	Inundation, soil moisture	6410	Used steps: 1, 3, 4, 5 (Jussila model, NDMI, NDWI, STR, SWIR1 (B11), NDPI,NDVI), 6 (bi-weekly,	2-weekly monitoring. Most relevant season: whole year Min, mean, max values of moisture indices.	Github page: https://github.com/TamaraKirin/Wetness-Variability-in-

					min, mean, max aggregation), 7, 8 (temporal interpolation), 9 (half year period),10	Number of inundated timesteps. % of constantly inundated, seasonally inundated and permanently dry areas	Zvecevo.git
SLU_SE	Aapa mires (n=30) in Västerbotten and Norrbotten counties.	2015-2025	Inundation, soil moisture	7310	Used steps: 1,2,3,5,6,9,10 NDWI, NDVI, Jussila model. In addition, drone data collection and from 3 mires, and subsequent water mask creation	Monthly monitoring during snow free vegetation season, May to September. Yearly min, mean, max values of moisture indices Number of inundated timesteps. Creation of water mask for calibration / validation of inundation index	
SLU_SE	Örsten, Västerbotten County (1 site)	July median MSAVI product for 2016-2025	Inter-annual changes in bare sand following a local restoration intervention	21XX	Used steps: 1, 2 (July), 3, 5 (MSAVI), 6 (monthly), 9, 10	Monthly monitoring. Relevant month: July. Median values per pixel. Pixel thresholding (MSAVI < 0.50 = bare sand). % of bare sand for July, monitored for each year	
BOZEN	Lago di Caldaro wetland area	01/01/2020 - 31/12/2024	Inundation, soil moisture	3140, 6410, 7210	Used steps: 1, 2, 3, 5 (Wiw, Jussila model, NDMI, NDWI, STR, SWIR1 (B11), NDPI), 6 (bi-weekly, monthly, min, mean, max aggregation), 7, 8 (linear interpolation), 9,10	2-weekly or monthly monitoring. Most relevant season: Spring - Autumn Yearly min, mean, max values of moisture indices. Between-year comparison of aggregated monthly mean values of monthly indices Number of inundated timesteps.	

Selecting spectral indicators

Several indices, bands and model predictions were used as indicators for habitat wetness, soil and vegetation moisture and water-cover. For dunes, Modified Soil-Adjusted Vegetation Index (MSAVI) was used with a threshold to separate bare sand. MSAVI suppresses soil background effects better than NDVI (Rouse et al. 1974) under sparse vegetation typical of dunes. The formula of indices, as well as their bio-physical relationship and link to the original paper are given in A1_Table 2.

A1_Table 2 Used indices in this subtask with their relation to bio-physical properties. B03, B8A and B11 are the green, near-infrared (NIR) and the first short-wave infrared bands (SWIR1) of Sentinel-2 L2A digital number (DN) values, respectively.

Index	Formula	Bio-physical attribute	Publication
Short-wave infrared 1 (SWIR1)	Band 11 of Sentinel-2 (B11)	Lower pixel values indicate higher soil (or vegetation) moisture of content.	(Sinergise, z.d.)
Transformed SWIR index1 (**STR1**)	((1-B11/10000)^2)/(2*B11/10000))		(Sadeghi et al., 2017)
Normalized difference moisture index (**NDMI**)	(B8A - B11)/(B8A + B11)	Higher pixel values indicate higher moisture content in vegetation.	(Gao, 1996)
Normalized difference water index (**NDWI**)	(B03 - B8A)/(B03 + B8A)	Higher pixel values indicate higher water content in water bodies.	(McFeeters, 1996)
(Modified) Normalized difference pond index (**NDPI**)	-NDPI = (B03 - B11)/(B03 + B11)	Higher pixel values indicate a higher chance of the pixel being a pond (based on aquatic vegetation, suspended materials and some types of soil).	(Lacaux et al., 2007)
Inundation	See inundation subtask report	Binary classification of inundated versus non-inundated pixels.	(Jussila et al., 2024)
			(Lefebvre et al., 2019)
Modified Soil-Adjusted Vegetation index (MSAVI)	(2 * B08 + 1 - sqrt((2 * B08 + 1)**2 - 8 * (B08 - B04))) / 2)	Values under 0.50 were classified as bare sand.	(Qi et al. 1994)

Filtering observations by quality or non-target areas (step 3)

Sentinel 2 L2A products have a **scene classification (SCL) layer** which assigns the Sentinel image data pixels into 12 classes using the machine learning model (Sen2Cor processor) of ESA (A1_Fig 1).

Label	Classification
0	NO_DATA
1	SATURATED_OR_DEFECTIVE
2	DARK_AREA_PIXELS
3	CLOUD_SHADOWS
4	VEGETATION
5	NOT_VEGETATED
6	WATER
7	UNCLASSIFIED
8	CLOUD_MEDIUM_PROBABILITY
9	CLOUD_HIGH_PROBABILITY
10	THIN_CIRRUS
11	SNOW

A1_Fig 1 The Scene Classification Values in L2A products off Sentinel 2.

For most cases, only pixels with labels with values 4, 5 or 6 (= target labels) were used for further analysis. Those scenes were used, in which these three labels provided a great majority of the total amount of pixels. Thresholds typically range between 95% and 80% for the different use cases provided by the partners (A1_Table 3).

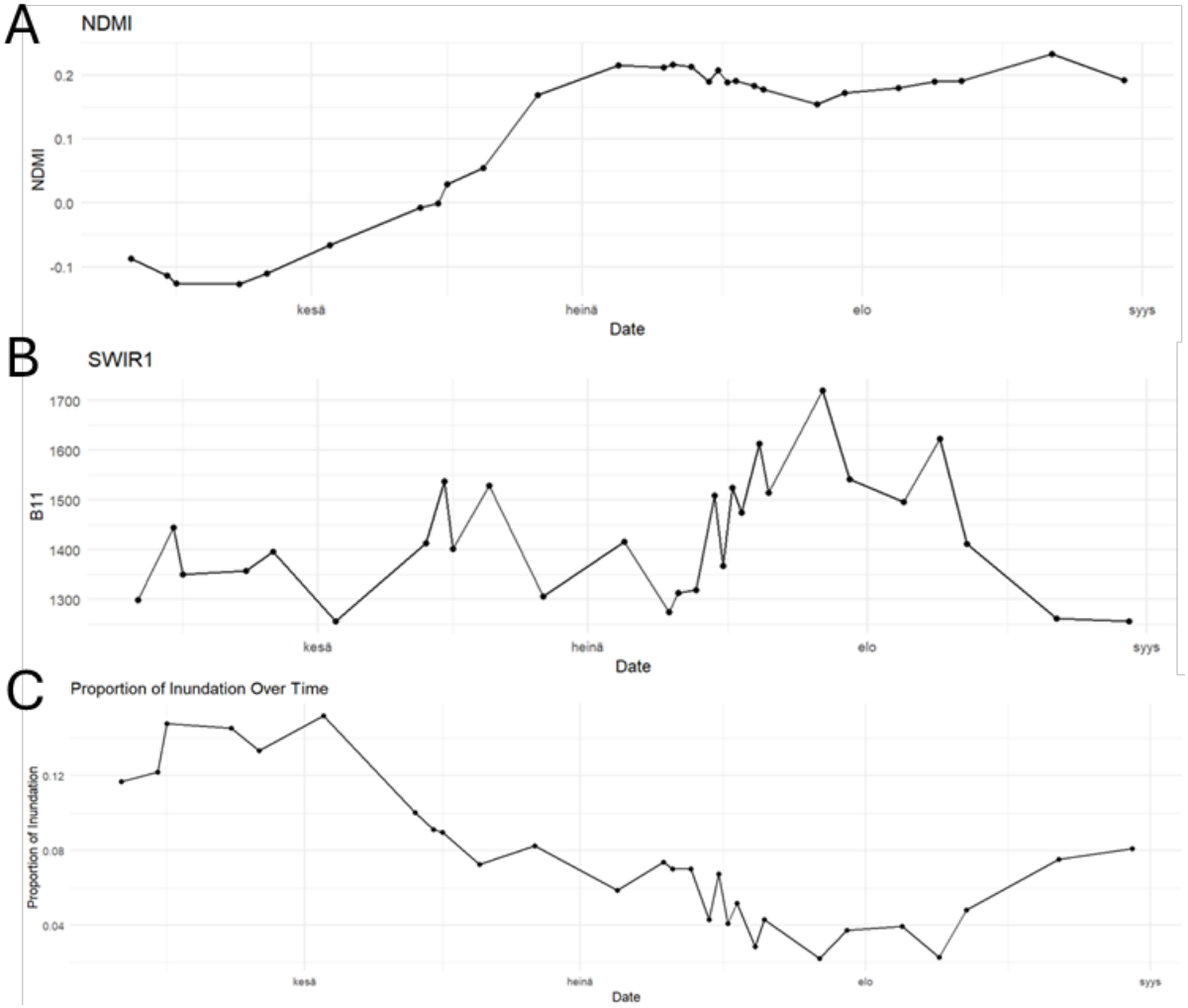
Within the used scene, non-target labels (clouds, snow, shadows, unclassified) were masked out. Water-covered pixels were sometimes problematically labeled as “unclassified” (label 7) in the SCL, which required modifications to masking in some pilot cases.

In our cases, we focused on collecting time series in natural areas. Further, we opted to mask out tree-covered areas, due to reported poor performance of many spectral indicators in situations where water surfaces occur under dense tree canopy (Burdun et al. 2020, Räsänen et al. 2022, Jussila et al. 2024). The same technical challenge was observed in the Inundation subtask of the pilot as well. For tree-masking, partners used either national data sources or the Pan-European canopy height dataset

(https://gee-community-catalog.org/projects/meta_trees/).

A1_Table 3 Effects of setting the threshold for the required target label percentage (%) on the amount (%) of available scenes. This example was performed for a 5 year period (01/01/2020 - 31/12/2024) for an area in Flanders (Webbekomsbroek, 1208,940 ha area).

Threshold (> %) of required target labels	95%	90%	80%
Number of available scenes	145	155	172
% of the total number of scenes (853 scenes in total)	17%	18%	20%



A1_Fig 2 A raw time series including all cloud-free observations plotted for A) a normalized difference moisture index (NDMI), B) a single band index and C) an aggregated model output (% of inundation). Remaining unfiltered atmospheric effects produce noise into time variables depending on single bands (B, C).

Occasional erroneous observations remained, even after filtering non-target labels and masking out trees, in the time series. These errors were caused by cloud shadows, atmospheric effects and thin clouds that were not detected by the Sen2Cor model of ESA. This kind of unfiltered low-quality observations in satellite image time series are rather common, and must be considered with respect to the uncertainties they may produce in the image interpretation and analysis. Normalized indices (e.g. NDMI A1_Fig 2A) are relatively robust against atmospheric errors and shadows, whereas single band indicators (e.g. SWIR1 [B11], A1_Fig 2B) are more susceptible to noise. Thus, handling single-band data, and classifications relying on single bands (e.g. inundation by Jussila's model, A1_Fig 2C), requires some caution. However, the temporal aggregation performed in the next step can help to smooth out most of the remaining noise (A1_Fig 4).

Temporal aggregation and gap-filling (steps 6-8)

The revisit time of Sentinel-2 satellites is 5 days and can be 2-3 days for some areas in Europe due to overlapping swaths. However, due to clouds, a large proportion of these raw observations were excluded from our applications (A1_Table 2), leading to sparse time series in which observations

are unevenly distributed in space and time. This can increase the risk of missing fast dynamics in the habitat, as was seen in the case of Croatia. The imbalance of the sparse time series is visualized in A1_Fig 3 and A1_Fig 4.

Optimally, observations should be distributed evenly across time for calculating comparable statistical indicators over larger time spans (seasonal, annual , 5-yearly, and other time slice delimited indicators). The imbalance creates uncertainty, especially in highly incomplete time steps. Gap filling techniques can be used to remedy incomplete time series with temporal gaps into more balanced, regular time series.

Possible approaches for gap filling are:

(1) combination of observations of different satellite imaging sensors (e.g. Landuyt et al., 2024; Neigh & Taylor, 2023),

(2) data prediction by active satellite imaging sensors (e.g. Van Tricht, 2023),

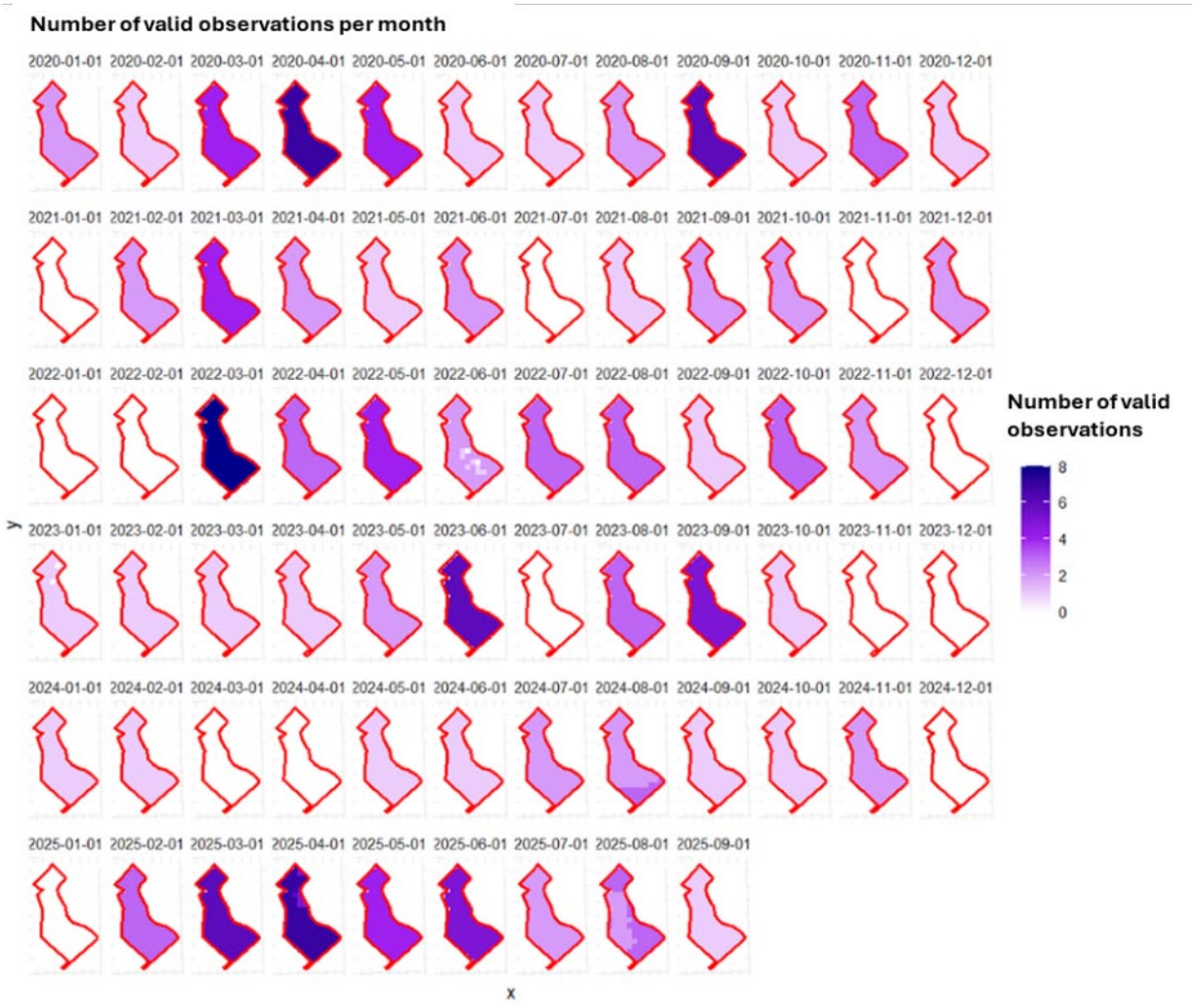
(3) cross-correlation of satellite observations with ground based sensor with high temporal coverage,

(4) temporal aggregation (compositing)

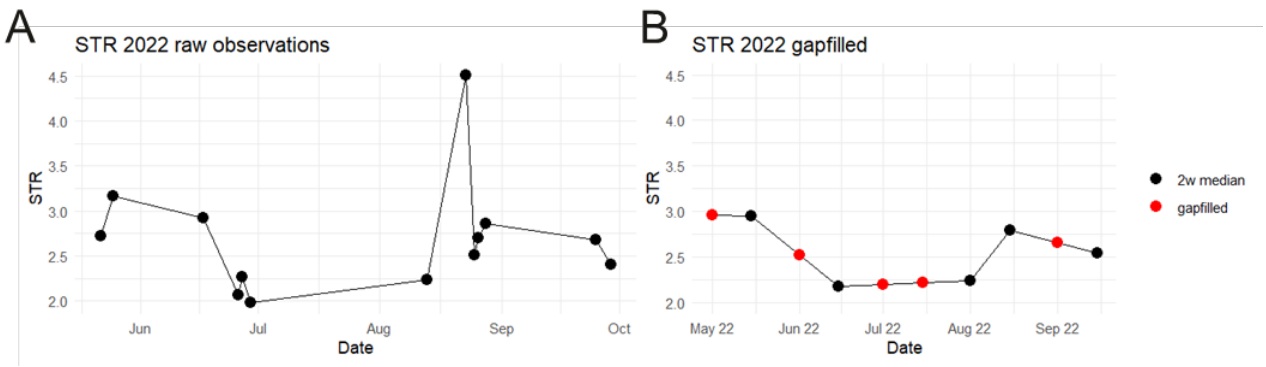
(5) spatial- and temporal interpolation (and/or smoothing) (e.g. Consoli et al., 2024)

All these techniques have their own benefits and challenges and would require further testing for habitat monitoring. Some basic versions of option 4 and 5 were examined during this subtask (link to datasets and code in A1_Table 1).

Yearly indicator statistics (mean, min, max) for trend analysis and maps are calculated from the aggregated and equidistant time series. Yearly minimum and maximum values , particularly with single band indicators, can be highly sensitive to outliers (originating from the signal or from the model). They are, however, useful for detecting “rare” events, like occasional inundation in very dry regions or drying in very wet regions. On average, aggregation to 2-weekly or monthly composites reduces the abundance of outliers, and is ecologically meaningful and sufficient time-scale in most habitats according to our spreadsheet survey (available for download at: https://doi.org/10.5281/zenodo.17712406).



A1_Fig 3 The problem of unbalanced (irregular) time series due to gaps. In a complete time series, the number of valid pixels should be 9 - 10 observations per month (revisit time of Sentinel-2 data for this region in Flanders is 3 days). The figure shows that this is far from true in most months. In a temporally complete monthly time series, we should have at least one observation for each pixel within each month. However, in some months (November, December, January, February and July) there are no valid observations present).



A1_Fig 4 STR-moisture index time series for one season, with A) all cloud free Sentinel-2 observations and B) bi-weekly aggregated and gapfilled observations. Observations are irregularly scattered across the season, and for half of the bi-weekly time steps observations are missing altogether and must be modeled with gapfilling. Linear

gapfilling was used for this case. Temporal aggregation also reduces noise from single erroneous observations (high spike in August).

Results

Habitat specific hydrology monitoring needs and limitations: survey results

The results of the spreadsheet survey on habitat specific monitoring needs and limitations are available for download at Zenodo (https://doi.org/10.5281/zenodo.17712406). The final table includes information for 12 wetland types, 18 grassland types, 1 freshwater type, and 1 forest type provided by seven partners, including the regions of Bolzano (Italy), Catalonia, Czech Republic, Denmark, Finland, Flanders (Belgium), and Slovakia.

According to the survey results, sufficient temporal resolution to detect ecologically relevant hydrological phenomena was weekly or bi-weekly for most habitat types. Even monthly information was considered useful, though higher temporal resolution is preferred for detecting finer temporal changes in variables such as length of inundated period. High temporal resolution monitoring was identified as necessary only for alluvial meadows of river valleys (6440), where very short-term flooding events can occur.

The experts also identified some limitations for RS-monitoring, depending on habitat type and region. Hydrology monitoring in some cases was expected to be very challenging due to tree-cover (7120 Degraded raised bogs, 9210D Bog woodlands) or due to very small size of typical habitat occurrences (for example, 7230 Alkaline fens in Bolzano and Finland). Habitat margins are typically tree-covered for multiple habitat types, but central areas in the majority of wetland and grassland types are usually open, or have only sparse tree-cover, allowing direct monitoring of land surface conditions with (optical) RS. Small-scale heterogeneity and vegetation cover in water areas were identified as possible limitations for multiple wetland habitats, which can negatively affect Sentinel-2-based water surface detection (see chapter for Inundation task).

Needs for specific hydrological metrics (season/yearly mean, min, max wetness) were somewhat similar across habitat types, as all the proposed metrics were considered relevant for most habitats. Maximum wetness of season was considered a slightly less interesting variable compared to minimum, which reflects the hydrological state at the driest week or month. Moisture indices were considered useful hydrology indicators in all habitats. Inundation monitoring was considered relevant in all wetland habitat types, and in 7 (out of 18) grassland habitat types. Overall, hydrological indicators were considered very useful for wetlands, and also interesting for many grassland types though not as highly relevant for habitat condition monitoring.

Inter-regional differences in monitoring needs were related to relevant seasons and condition interpretation of hydrology indicators. The snow-free period from spring to autumn was largely agreed as the most relevant season for monitoring, but naturally the length of that season varies depending on region. For example, in Flanders even the winter months are interesting for monitoring, whereas in northern Finland and Sweden the snow-cover can last until May or even early July. Habitat-type specific differences in preferred indicators and their interpretation arise from the ecological differences between the habitat: for example, prolonged inundation is interpreted as a potentially negative factor for ecological condition in alkaline fens (7230) and *Cladium mariscus*

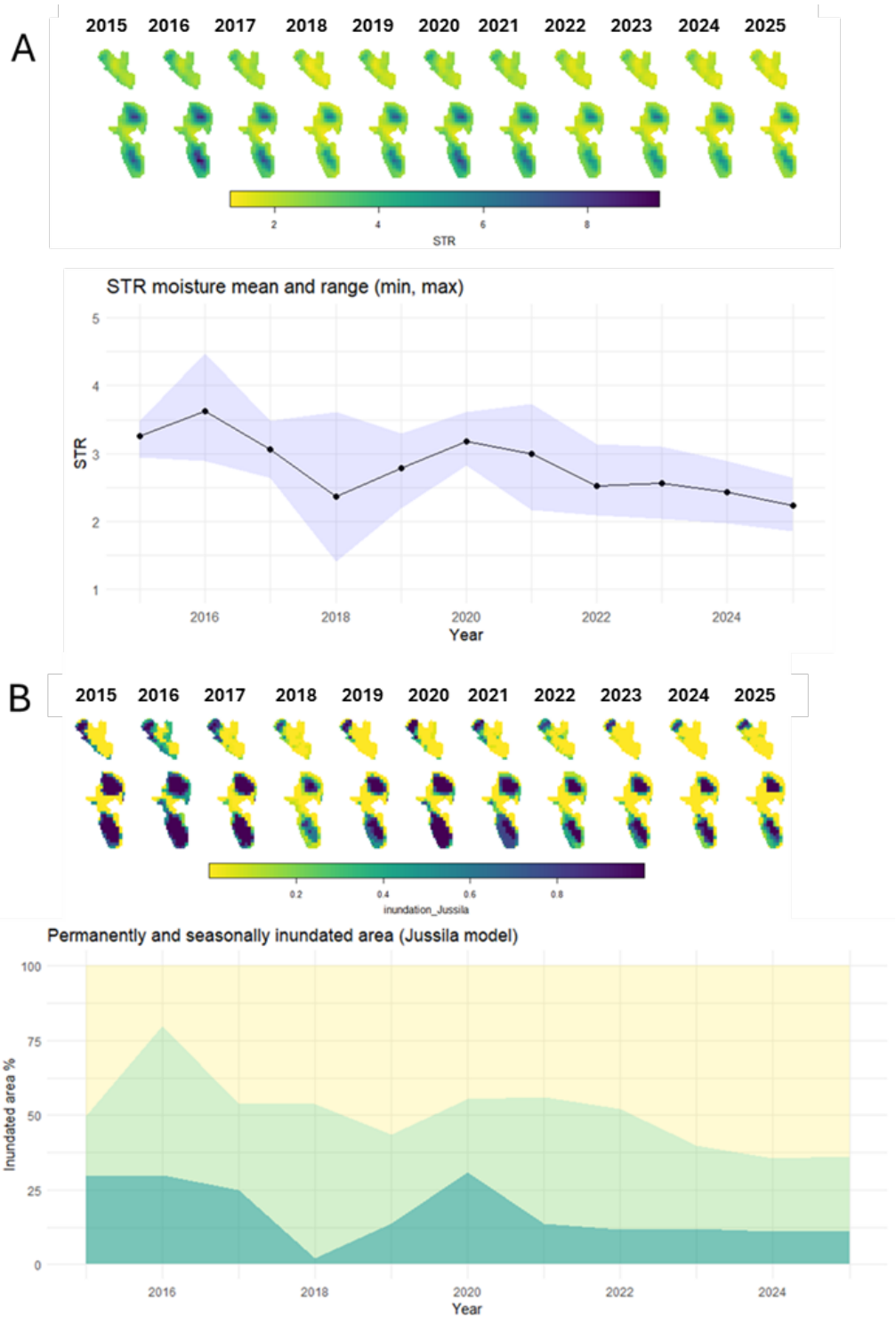
stands (7210), whereas for aapa mires (7310) the presence of both seasonally and constantly inundated areas is a vital part of habitat condition.

Indicator outputs

With the standardized processing pipeline, (multi)yearly indicator time series were processed automatically and their results could be easily visualised for the study sites. Graphs provide useful information at the plot level (e.g. patch metrics such as percentages, fragmentation, and edges), while raster time series provide pixel level information (e.g. distribution). Temporal dynamics can indicate trends (disturbances, recovery, losses, gains, stability, restoration effects).

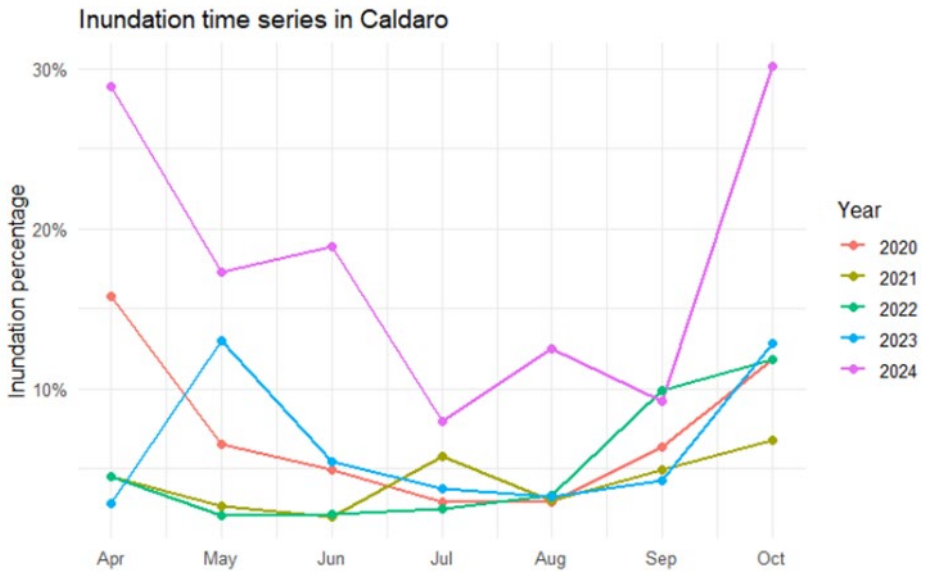
The indicators and their usage in the subtask included, for example:

- yearly mean, min, max of moisture indicators (e.g. STR index) with high values indicating higher soil moisture (A1_Fig 4A).
- yearly inundation statistics: amount of seasonally and permanently wet surface (percentage of time under watercover) (A1_Fig 4B).
- interannual variability in seasonal inundation regime (A1_Fig 6)
- summary tables about the timing and duration of inundation automatically created per year by setting simple rule based thresholds on the percentage inundation on parcel level (A1_Fig 7).
- multiyearly indicator change maps to provide useful spatial information on changes in trends (e.g. the decrease in inundation) (A1_Fig 8).
- evaluation of dune restoration effects in combination with high spatial resolution ortho imagery (A1_Fig 11).

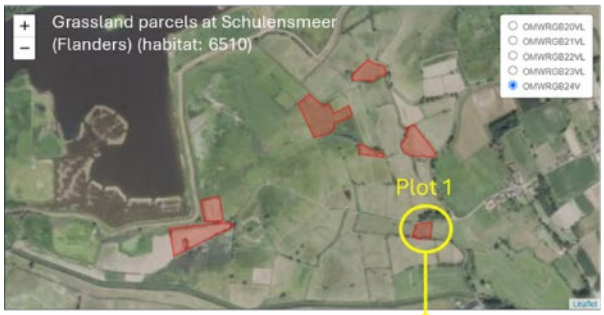


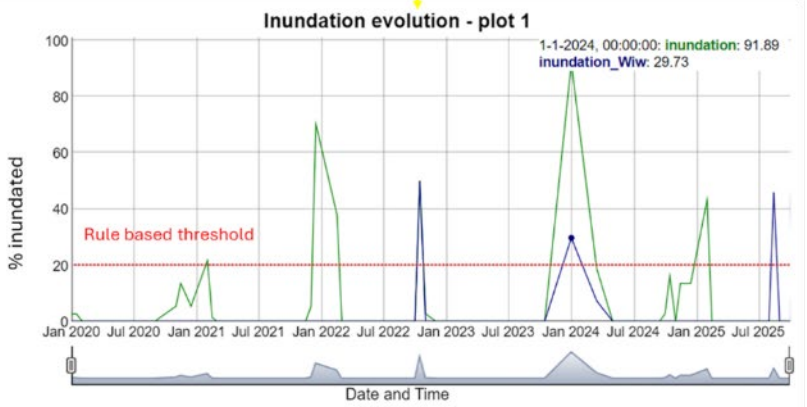
A1_Fig 5 Graph and raster visualisation of the soil moisture index STR (5A) showing the relative soil moisture (averaged per year) within individual habitat plots, and the graph summary of the plot mean, min and max values per year. 5B: Raster time series of permanently dry (yellow pixels) to permanently inundated (dark blue) pixels per year. Intermediate values are seasonally inundated pixels, with the relative frequency of inundation indicated by the color palette of the legend. The proportions of permanently dry (yellow), seasonally inundated (light green) and permanently inundated (darker green/blue) area per year are visualized in the bottom graph. Example from an aapamire in Finland.

Appendix 1. Indicators for Habitat condition monitoring



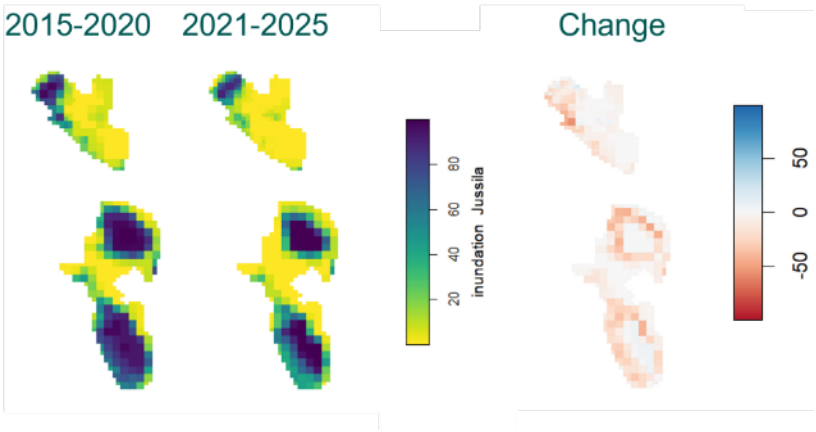
A1_Fig 6 Visualisation of seasonal changes of inundated area within a habitat, comparing five different years. The year 2024 had high precipitation and shows higher-than-normal extent of inundation throughout the season. Example from Bolzano, Italy.





Year	Classification model	Start	End
2020	Jussila (inundation)	/	/
2020	Wiw (inundation_Wiw)	/	/
2021	Jussila	1: 2021-02-01 2: 2021-12-15	1: 2021-02-01 2: 2021-12-15
2021	Wiw	/	/
2022	Jussila	1: 2022-01-01 2: 2022-10-15	1: 2022-02-15 2: 2022-10-15
2022	Wiw	2022-10-15	2022-10-15
2023	Jussila	2023-11-01	2023-12-15
2023	Wiw	2023-12-15	2023-12-15
2024	Jussila	2024-01-01	2024-03-01
2024	Wiw	2024-01-01	2024-02-01
2025	Jussila	2025-01-01	2025-02-01
2025	Wiw	2025-08-15	2025-08-15

A1_Fig 7 Example of an inundation monitoring scheme for 6510 grasslands in Flanders. Start and ending of inundation season(s) per year are determined by a simple rule based threshold. In this example, we consider the area as inundated if more than 20% of the parcel is inundated. This threshold is however flexible and used here mainly for illustrative purposes.



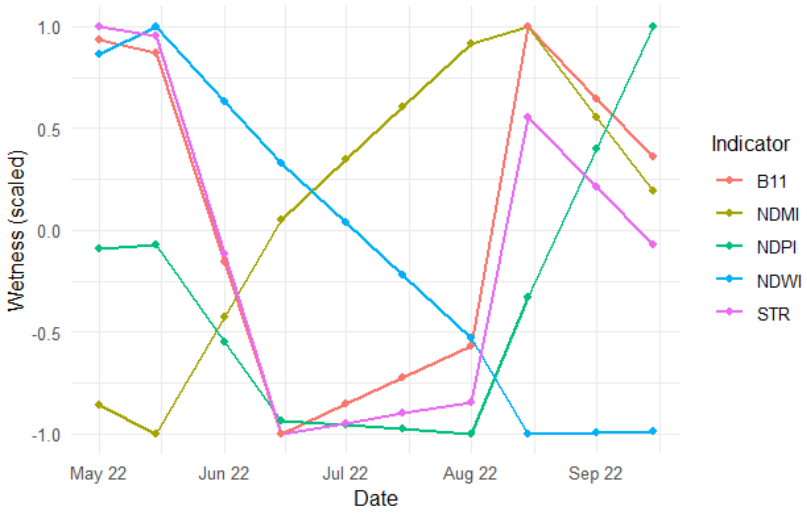
A1_Fig 8 Inundation seasonality rasters aggregated to two imaginary five year monitoring periods (2015-2020, and 2021-2025), showing the average inundated time (inundated % of time within the snowless season). The change map in the right shows the difference between the two periods, showing decrease of inundation in the edges of wet areas. Example from a Finnish aapa mire.

The results can be fundamentally affected by the selection of the spectral indicator. For wetness, the different indices and moisture-related bands show differing temporal patterns especially seasonally (A1_Fig 9). Differences occurred in the yearly trends as well.

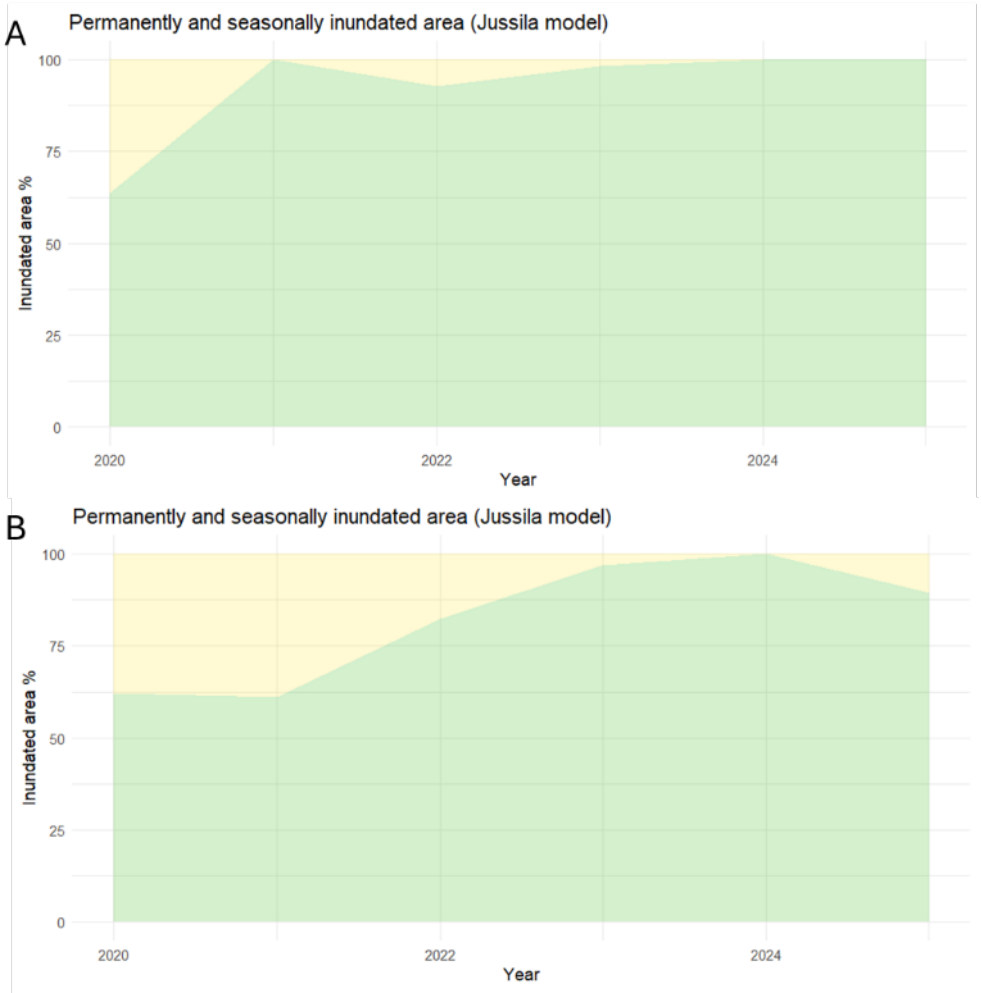
Also, sensitivity to changes in (pre)processing parameters was observed (A1_Fig 10). In order to compare and interprete results of habitat quality monitoring, pre-processing approaches should be harmonized for similar habitats. The example below shows the effect of a single parameter (sensitivity analysis) on determining the percentage of dry, seasonally inundated and permanently inundated area for a study site in Flanders. For both subplots, the same preprocessing strategy was used (same raw timeseries, temporally aggregated with median function, temporally interpolated (linear), min/mean/max function for yearly indicators). Only the output resolution of the temporal aggregation is different between the two figures (2 weekly for Figure 10A, monthly for A1_Fig 10B).

A1_Fig 9: Comparison of the different wetness indicators/indices shows highly varying, even opposite, seasonal patterns. Higher values indicate higher wetness level. Example from Finland, where aapa mires are usually flooded in spring, decrease in wetness during summer, and get wetter again in autumn.

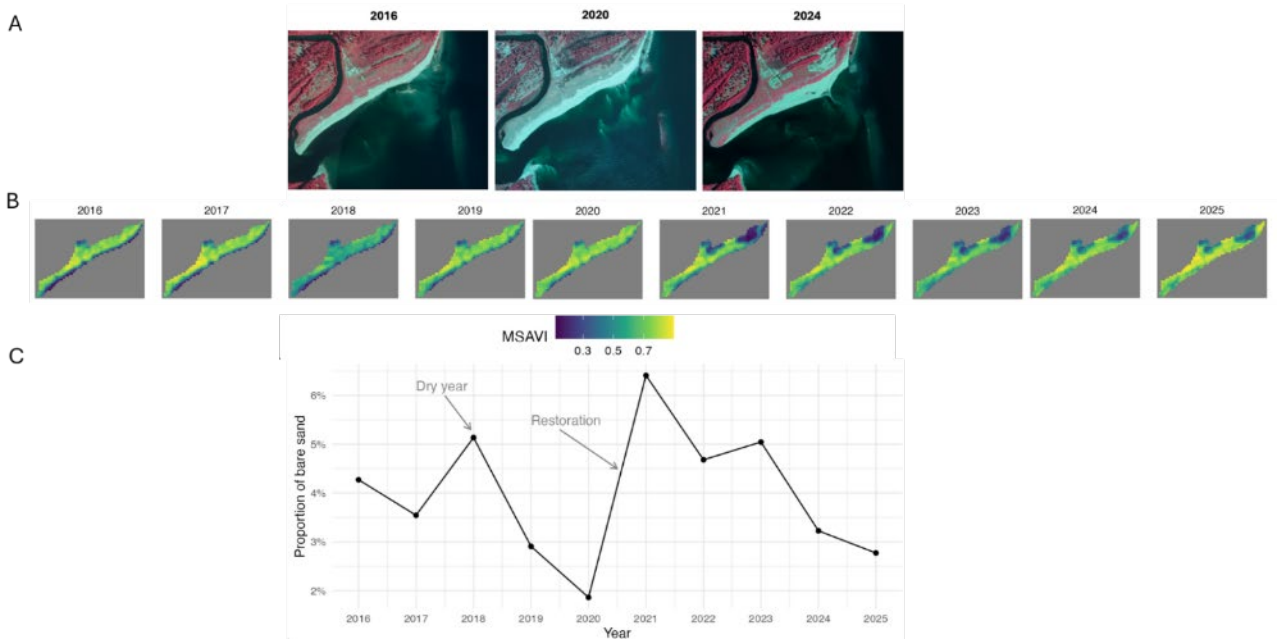
Appendix 1. Indicators for Habitat condition monitoring



A1_Fig 9 Comparison of the different wetness indicators/indices



A1_Fig 10 Sensitivity to one parameter in the time series (pre)processing pipeline. 11A: is 2 weekly median temporal aggregation, while 11B: is monthly median temporal aggregation of predicted inundation by the Jussila model. Both graphs indicate an increase of seasonally inundated area, but the timing and magnitude is highly different between the two graphs. Example from Flanders, Belgium.



A1_Fig 11 Example of dune restoration monitoring in Sweden. 11A: False color infrared aerial images from 2016, 2020, and 2024. Large restoration patches are clearly visible in 2024 in the eastern part of the dune field. 11B: July median Modified Soil Adjusted Vegetation Index (MSAVI) for each year. Dark blue tones indicate bare or sparsely vegetated sand, while light green to yellow indicates denser vegetation. Restoration patches are evident from 2021–2025. 11C: Sentinel-2 L2A time series of bare sand cover, using MSAVI < 0.50 as the classification threshold. The exceptionally dry summer of 2018 stands out with reduced vegetation cover in July.

Discussion

This pilot explored the path from raw satellite observations to analysis-ready yearly time series of meaningful habitat condition indicators. The hydrology case exemplified the process with a special focus on European-wide harmonisation, demonstrating

a) The role of remote sensing (RS) -based methods in providing unprecedented, upscalable information on habitat hydrology and its spatial and temporal variation

b) The creation of easy-to-apply processing pipelines for automated indicator time series generation, upscalable to a considerable habitat site sample (tens of sites tested in this pilot, feasible to even hundreds of sites, depending on their size)

c) The various decision points along the pipeline that affect the results and, thus, the harmonisation of monitoring

d) The importance of a strong scientific foundation and mapping the hydrology-related monitoring needs and expected limitations across habitat types and regions

Demonstration of harmonised condition monitoring

We successfully applied Sentinel-2 satellite data to produce and visualise simple yearly indicator sets on hydrology for our test sites across multiple EU regions and habitat types. Time series of yearly average of a moisture index, together with variation from the driest to the wettest period of the year, reflect the overall moisture levels and variations in any open habitat type, and can be used to

detect the occurrence and assess the severity of extreme events such as floods and droughts. For the wetland and mire habitats (and some grassland habitat types) in which inundation-regime or presence of open water is relevant information for habitat condition, we showcased the yearly duration of the inundated period, and yearly area of permanent and seasonal inundation as potential indicators.

Based on our pilot survey, the key ecohydrological information needs - such as seasonal wetness metrics, minimum moisture levels, and inundation dynamics - are largely consistent across habitats and regions. This supports the feasibility of monitoring them with harmonised RS-based indicators such as those produced in this pilot. Some specialities arise from differing season lengths. Winter months for example are not relevant for all due to snow-cover, and in some habitats late autumn outside growing season was not considered ecologically highly relevant for monitoring. Yearly average wetness and other indicators are considered most useful when applied only on the (regionally varying) relevant seasons. When finally interpreting the condition status with RS-indicators, naturally also the favourable reference levels to wetness vary between habitat types and possibly regions as well.

These kinds of new, satellite-based indicators provide not only a possibility for harmonised monitoring, but a fundamentally novel monitoring approach for hydrological condition monitoring in wetlands and grasslands. As seen from the time series, interannual and within-year variation in habitat wetness is high, which emphasizes the benefits of continuous monitoring and long-term trend analysis. In addition to a temporally extensive view, the approach is also spatially scalable, which means that the same indicator data can easily be generated for any geographical level, to show temporal trends at the individual site level as well as across an entire e region. A combined use of time series plots and raster maps provides a powerful dashboard for monitoring from different perspectives - whether the focus is on localising/detecting changes of any size (pixel-level phenomena) within a site, or on general site-level trends through time. Multilevel approaches allow effective monitoring of habitat type's regional condition generally, as well as spatially explicit detection of degradation, restoration effects, recovery or climate change impacts.

Technical considerations for upscaling indicator production

In this subtask, we demonstrated a reproducible and easy-to-apply R-based processing pipeline for calculating a set of indicators. The same scripts, customised to local data and sites, were applied by multiple pilot partners. The approach (openEO for retrieving datacube followed by local analysis on computer) enables independent production of tailored multi-year time series on habitat condition indicators for tens of sites. For example, the ten-year time series for the Finnish sites (n=20) were retrieved via openEO with 300 credits (of monthly 10 000 free credits), effectively with no processing costs or special requirements for computing power or data storage (1 GB). The approach could be applied to a sample-based monitoring of habitat condition for an entire country, as well as research on method development and validation.

Transferring the whole processing pipeline (instead of only retrieving image time series) to a cloud platform (such as openEO or SentinelHub APIs) requires extra steps of development with skills on cloud computing, but is a prerequisite for extensive larger scale application for operational biodiversity monitoring. Examples, such as a mire wetness dataset from Finland (https://ckan.ymparisto.fi/fi/dataset/vetiset-suonpinnat), show that country-level multi-year and even monthly indicator products of a single variable are feasible nationally with relatively low cloud

computing costs. However, continental scale analysis of multiple variables and for long monitoring periods, in an operative manner, naturally requires more technical and financial resources. The Global Surface Water Explorer (GSWE, https://global-surface-water.appspot.com/map) is a good example of global-level products.

The large data volumes and computational demands of satellite time series analysis requires cloud computing and high-performance computing (HPC) infrastructures that can handle petabyte-scale data and perform continuous analyses at continental scale. Until today, these costs are quite low for non-commercial and research use through platforms like **Google Earth Engine (GEE)**, and more recently, the community based **OpenEO,** which is part of the **Copernicus Data Space Ecosystem**.

It is highly important that these costs will remain low in the future. For GEE, uncertainty stems from the fact that the platform is not open source, and its access policies or pricing models may change over time. OpenEO is open source, but long-term sustainability and continued community or institutional support must be guaranteed, to minimize uncertainty regarding future costs. We therefore hope that this community will remain active in the future, as it will be key for the successful monitoring of habitats through RS on an European level.

Limitations and uncertainties

The information path from satellite sensors to analysis-ready indicator products contains many steps, each posing challenges and introducing sources of uncertainty, which lead to limitations in the use of these products. Some of these limitations require merely cautiousness in information processing in order to produce high-quality products, and some would require further research efforts and development.

First of all, **validation of spectral indicators and the time series** is of critical importance. In this pilot, we applied several different moisture indicators (NDMI, STR, SWIR, NDWI), which gave partly very **contradictory views on wetness** particularly in time, seasonally and between years. We did not have the resources for extensive ground truth measurement campaigns within this pilot, and were thus unable to properly assess the cause of these differences. Some indices might perform better than others, or they might reflect different aspects of wetness (moist green vegetation vs. soil water content). Extensive **ground truth data** is necessary to support an informed selection of suitable indices for a specific habitat type. This ground truth validation should cover the whole monitoring season and drought periods if possible, to properly account for sensitivity of many moisture index signals on vegetation phenology. Continuous ground-based measurements might be a necessary support for remote sensing monitoring programs to ensure reliability in a changing environment. Moreover, the **selection of a spectral indicator can significantly alter the estimated condition,** and even the indicator trendlines, and is thus **important for harmonisation** as well.

Secondly, the (optical) **satellite-based indicators have characteristics that can limit their use in certain habitat types**. Tree-cover and thick reedbeds were identified as one main limitation, that based on our observations causes misclassification of inundation (see the Inundation subtask report), and affects spectral soil moisture signals as well (Burdun et al. 2020, Räsänen et al. 2022). Based on our survey, many wetland and grassland habitats can be **partly tree-covered** and require masking out the tree-covered parts. The few habitat types with dominant tree-cover require a dedicated method development, because our methods cannot guarantee reliable results for them. For some habitat types, increasing tree-cover is one indicator of degradation, and simultaneous

monitoring of wetness and tree-cover requires caution. **Reedbeds** and other dense, locally partially submerged vegetation is rather common in many wetlands types, and would need targeted ground truth data collection and dedicated method development/refinement. **The pixel size** of freely available satellite imagery (e.g. Sentinel-2) poses another limitation. It is a challenge for inundation detection in spatially heterogeneous habitats (see the Inundation report), and also makes monitoring of very small habitat patches challenging or practically impossible, possibly ruling out some habitat types from satellite-based monitoring. Many high-resolution satellite products **do not include** the **SWIR bands** necessary for calculating the hydrology indicators tested in this pilot.

Thirdly, we need to consider the common satellite observation-related sources of uncertainty: atmospheric interference and gaps in time series. **Transparency of processing pipelines** and reproducibility of workflows are critical for trust and uptake in policy and management. Differences in pre-processing (e.g. atmospheric correction, cloud masking), temporal compositing, and **gap-filling techniques** can introduce significant variability and uncertainty into the derived indicators. The impact of these methodological choices on the uncertainty of time series analyses should be explicitly quantified and documented. Moreover, ensuring **spatial and temporal completeness** of time series data is essential for reliable trend estimation, particularly in regions with frequent cloud cover or strong seasonality. Communication of uncertainty in indicator products is highly recommended, and becomes particularly important and challenging when the level of aggregation increases. For a single site and a single satellite observation, errors can be still detected visually, but these errors become increasingly “invisible” when moving to aggregation of multiple observations into yearly indicators, or aggregating indicators to include multiple sites over large areas. Regarding gapfilling uncertainties, for example, one can report completeness of the raw data behind the time series spatially (the number of valid pixels within an area) and temporally (the proportion of valid time steps).

Harmonizing habitat monitoring across Europe will require **common standards for indicator definitions, data processing, and validation protocols**. Establishing shared reference datasets, open-source processing chains, and cross-country calibration frameworks could enhance interoperability and comparability. European-scale initiatives, such as the Copernicus Land Monitoring Services and EIONET, provide a strong foundation for such coordination.

From indicators to condition assessment

Finally, further research and harmonisation needs are likely to arise from practical application of the indicator variables to answer policy-relevant questions. RS-based indicators, such as produced in this pilot, are indicative of ecologically important biophysical phenomena, and can be used to detect change and trends over time periods. However, what remains an unanswered question, is the final interpretation of the habitat condition status based on the RS indicators. The practical application of condition assessment relies on standardizing reference values so they can be directly aligned with (or converted to) the outputs from the RS-based indicators. For hydrology indicators, as an example, these could be either historical reference levels (from satellite era) for the site, or habitat type specific reference levels, such as explicit threshold of sufficient proportion of water cover, or favourable range of moisture index values. Each of these reference approaches is expected to have its own methodological challenges - such as limited historical perspective, or difficulty to define thresholds if there is large natural variability within a habitat type. The availability of reliable habitat maps (possibly RS-based) is another prerequisite for regional RS-based condition assessment.

Solutions for these limitations are needed before RS-based indicators can be applied to fulfill policy needs such as monitoring the condition of the Annex I habitats (Habitats Directive), or assessing restoration success (Nature Restoration Directive). Further development and harmonisation efforts should be directed on application of RS-based products for condition assessment in order to get the full benefits of the RS capabilities for biodiversity monitoring.

References

Burdun, I., Bechtold, M., Sagris, V., Lohila, A., Humphreys, E., Desai, A. R., Nilsson, M. B., De Lannoy, G., & Mander, Ü. (2020). Satellite Determination of Peatland Water Table Temporal Dynamics by Localizing Representative Pixels of A SWIR-Based Moisture Index. *Remote Sensing*, *12*(18), 2936. https://doi.org/10.3390/rs12182936

Consoli, D., Parente, L., Simoes, R., Şahin, M., Tian, X., Witjes, M., Sloat, L., & Hengl, T. (2024). A computational framework for processing time-series of earth observation data based on discrete convolution: Global-scale historical Landsat cloud-free aggregates at 30 m spatial resolution. *PeerJ*, *12*, e18585. https://doi.org/10.7717/peerj.18585

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257-266. https://doi.org/10.1016/S0034-4257(96)00067-3

Jussila, Tytti, Risto K. Heikkinen, Saku Anttila, e.a. 'Quantifying Wetness Variability in Aapa Mires with Sentinel-2: Towards Improved Monitoring of an EU Priority Habitat'. *Remote Sensing in Ecology and Conservation* 10, nr. 2 (2024): 172-87. https://doi.org/10.1002/rse2.363.

Landuyt, L., Ivashkovych, X., & Van Achteren, T. (2024). Consistent Surface Water Monitoring by Fusing Sentinel-1 and -2 Through Convolutional Neural Networks. *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 4701-4705. https://doi.org/10.1109/IGARSS53475.2024.10640419

Lefebvre, Gaëtan, Aurélie Davranche, Loïc Willm, e.a. 'Introducing WIW for Detecting the Presence of Water in Wetlands with Landsat and Sentinel Satellites'. *Remote Sensing* 11, nr. 19 (2019): 2210. https://doi.org/10.3390/rs11192210.

Lacaux, J. P., Tourre, Y. M., Vignolles, C., Ndione, J. A., & Lafaye, M. (2007). Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. *Remote Sensing of Environment*, *106*(1), 66-74. https://doi.org/10.1016/j.rse.2006.07.012

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

Neigh, C., & Taylor, M. P. (2023, August 30). *Harmonized Landsat Sentinel-2 – Harmonized Landsat Sentinel 2*. NASA. https://hls.gsfc.nasa.gov/

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index (MSAVI). Remote Sensing of Environment, 48(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

Räsänen, A., Tolvanen, A., Kareksela, S., 2022. Monitoring peatland water table depth with optical and radar satellite imagery. Int. J. Appl. Earth Obs. Geoinformation 112, 102866. https://doi.org/10.1016/j.jag.2022.102866.

Sadeghi, M., Babaeian, E., Tuller, M., & Jones, S. B. (2017). The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. *Remote Sensing of Environment*, *198*, 52-68. https://doi.org/10.1016/j.rse.2017.05.041

Sinergise, S.-H. by. (z.d.). *Sentinel-2 Bands*. Sentinel Hub Custom Scripts. Accessed on 14 November 2025 on https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/bands/

Van Tricht. (2023, October 16). *New CropSAR service provides Cloud-Free Time Series at Pixel Level*. VITO Remote Sensing. https://blog.vito.be/remotesensing/cropsar2023

Annex

Survey: Hydrological information needs and limitations for habitat RS-monitoring. https://doi.org/10.5281/zenodo.17712406

Appendix 2. Inundation mapping

Partners involved:

- Stien Heremans stien.heremans@inbo.be (Research Institute for Nature and Forest - INBO)
- Jussila Tytti tytti.jussila@syke.fi (Suomen ympäristökeskus - Syke - Finnish Environment Institute)
- Verbesselt Sebastiaan sebastiaan.verbesselt@inbo.be (Research Institute for Nature and Forest - INBO)
- Bjärhall Albin albin.bjaerhall@eurac.edu (Eurac Research)
- Rataj Jakub jakub.rataj@aopk.gov.cz (Nature Conservation Agency of the Czech Republic - AOPK ČR)
- Leština Dan dan.lestina@aopk.gov.cz (Nature Conservation Agency of the Czech Republic - AOPK ČR)
- Ježek Vít vit.jezek@aopk.gov.cz (Nature Conservation Agency of the Czech Republic - AOPK ČR)
- Gardfjell Hans hans.gardfjell@slu.se (Sveriges Lantbruksuniversitet - SLU)
- Adler Sven sven.adler@slu.se (Sveriges Lantbruksuniversitet - SLU)
- Mária Šibíková maria.sibikova@savba.sk (Slovak Academy of Sciences - SAS)
- Jozef Šibík jozef.sibik@savba.sk (Slovak Academy of Sciences - SAS)
- Martin Kelko martin2kelko@gmail.com (Slovak Academy of Sciences - SAS)
- Marek Súľovský sulovsky.mark@gmail.com (Slovak Academy of Sciences - SAS)
- Tamara Kirin tamara.kirin@mzozt.hr (Institute for Environment and Nature, Ministry of Environmental Protection and Green Transition in Croatia)

Background

Inundation (or flooding) refers to the submersion of land by water which is a very important driver and indicator of ecosystem (and habitat) type, composition and quality. The presence of water, whether permanent or seasonal, creates a unique environment for several water-dependent species. The monitoring of inundation is especially important in wetland ecosystems, but is also relevant in other ecosystems such as wet grasslands and alluvial forests. The inundation regime can even serve as a condition indicator for habitat types where (seasonal) flooding is a characteristic part of the habitat's ecohydrology.

Field-based monitoring of inundation is extremely time-consuming, because inundation may be temporally dynamic as well as spatially variable over small distances. Therefore, remote sensing provides a very promising technique that allows the automatisation of repeated measurements over large spatial scales at a relatively detailed resolution.

The Sentinel-2 mission from the EU's Copernicus programme is particularly well-suited for inundation monitoring, as its twin satellites provide high-resolution (10m) optical imagery with a rapid five-day revisit time, crucial for capturing dynamic events. Furthermore, its multispectral bands in the Near-Infrared (NIR) and Short-Wave Infrared (SWIR) regions are highly effective at detecting water, which strongly absorbs radiation at these wavelengths, allowing for a clear differentiation of wet areas from dry ones.

However, inundation in wetlands and grasslands presents a special challenge for RS compared to the usual cases of open water body detection, since the water layer in these habitats can be extremely shallow (and with varying bottom surface properties) and partly or totally obscured by vegetation. From an ecological and ecosystem functionality perspective, already a centimeter of water and consequent water saturation of the root zone can be meaningful for water-cover monitoring. Moreover, the wetland and grassland habitats are often structurally heterogeneous and image pixels involve a mixture of water and non-water surfaces, affecting the performance of water detection methods.

To address these problems, several models have been proposed for an automatic binary mapping of inundated water-featured pixels versus dry pixels, particularly for wetland ecosystems. The overall objective of this subtask was to test and validate two Sentinel-2-based classification models for detecting inundation in open wet ecosystems with no tree cover across different regions in Europe. We tested two simple models that were developed in Europe, and that are easy to upscale to the different wetland (and wet grassland) habitats across the continent: (i) the model introduced by Jussila et al. (2024) and (ii) the Water in Wetlands (WiW) model developed by Lefebre et al. (2019).

Some specific objectives that we addressed are:

1. Developing readily applicable but efficient approaches for validation, including reference data collection using aerial imagery
2. Evaluating the models' performance and scalability to heterogeneous environments across Europe, with a special focus on spatial fragmentation and temporal dynamics.
3. Validating the use of super-resolution images as a means of addressing the spatial heterogeneity challenge.

Additionally, we looked at the more general approaches for water detection (water indices, GSWE map product) and their performance in the context of wetland and grassland habitats.

Recommendations

In this subtask, we tested two models for the mapping of inundated areas in wetlands and grasslands across Europe. In general, we found a high to very high classification accuracy (measured as the macro-F1 score) for all the tested models across different study sites in Europe where the validation was performed.The challenges for water detection varied across habitat types and regions, and include the presence of thick reedbeds or ambiguous semi-inundation (aapa mire flarks) that are harder to detect and classify correctly (see Jussila et al. 2024)than well-delineated open water bodies.

The impact of pixel-level mixtures on model performance is complex, although we often see a decrease in performance as the mixture level increases. However, there is a good probability for inundation to be detected as long as the majority of the pixel area is water-covered.

It can be rather straightforward to provide the inundation models studied in this subtask or the derived maps on a cloud platform across the EU, preferably through the Copernicus Data Space Ecosystem (CDSE), one of the back-ends of OpenEO. Here of course the cost of processing has to be taken into account.

It can be possibly interesting to provide the super-resolution versions as well. However, this would require much more processing, so we recommend offering this as an on-demand service.

The number of participating countries as well as the number of validation sites within each country was limited in the frame of this project. Therefore, we do recommend a more structured validation approach spanning the entire EU territory. This could be supervised by an EU institution like ESA or EEA. We have developed some specific recommendations for this validation, and they are listed in the next subsection.

Recommendations for EU-wide validation of inundation models

Aerial images enable the effective collection of large validation samples across broad regions. Open water bodies are easy to delineate from these images, making accuracy assessments for their detection feasible. The detection rate for open water was very good with the tested models and water indices.

However, wetland habitats also exhibit less evident forms of inundation that are meaningful for ecosystem functioning: thin layers of water above ground, inundation under high vegetation, or watery peat and moss surfaces in northern mires. Each of these creates ecologically special conditions despite the shallow water depth. When setting habitat-specific goals for inundation monitoring, it is important to determine whether the aim is to monitor open water or also these less evident types of inundation. The validation method should then be selected accordingly, and the reference data labeling should reflect these goals.

For these less evident inundation conditions, model performance is harder to evaluate. We find that validations based only on aerial images are risky because the interpretation of these areas is inherently difficult and subjective. When the validation data itself is of low-quality, it becomes impossible to conduct robust assessments of the true performance of remote sensing (RS) methods. Further work is clearly needed to validate water detection performance in these challenging environments.

For validation approaches in these areas of uncertainty, we highly recommend combining aerial or drone imagery with targeted field visits. Increasing the amount of high-quality validation data in these various areas is essential for improving high-quality inundation monitoring. In addition to the uncertainty areas described above, it would be beneficial to broaden validation across different seasons (to capture the effect of vegetation phenology on water detection) and weather conditions, including drought periods. Moreover, a method's ability to reflect changes should be validated with multitemporal reference data to ensure its suitability for monitoring.

Based on assessments of this subtask, the models trained specifically for wetlands showed promise even for less evident inundation. However, they have clear limitations. For instance, tree-masking is necessary for these models. Additionally, detecting inundation under reedbeds proved particularly challenging—both for labeling reference data and for the RS models—and requires further research. Moreover, as was demonstrated in Croatia, the minimum mapping unit for detection of inundated areas is 1 Sentinel-2 pixel. Wetland habitats smaller than this detection limit are not picked up by the models.

We strongly recommend an EU-wide validation of these (and other) inundation models. Such an extensive validation enterprise should be performed on a cloud-based platform like openEO as a part of the Copernicus Data Space Ecosystem. However, this endeavour should have a clearly defined scope, with detailed and systematic instructions regarding the minimal mapping units and the masking of non-open areas (e.g. trees and buildings) out from the habitat patches.

Work description

Methods and data

Overview of the validation sites

A2_Table 1 gives an overview of habitat types and the number of study sites used for validating the inundation models.

A2_Table 1 Overview of the habitat type(s) and number of study sites for validation of the inundation models in each region of the EU involved in this inundation subtask

Region	Habitat type	Nr of sites
Finland	Aapa mires, corresponding to the Annex 1 habitat 7130	7
Flanders	River floodplains (seasonally inundated) that contain Annex I (wetland and grassland) habitats as well as regionally important biotopes	4
Slovakia	River floodplains, mostly Annex I habitat 3150*	1
Czechia	River floodplains, mainly corresponding to Annex I habitat 3150*	2
Bolzano	Calcareous wetland area, containing the Annex I habitats 3140, 6410, and 7210	1
Sweden	Aapa mires, corresponding to the Annex 1 habitat 7130	3
Croatia	Mediterranean temporary ponds (Annex I habitat 3170)	6

*Natural eutrophic lakes with Magnopotamion or Hydrocharition -type vegetation

For Finland, six representatives of the Natura2000 aapa mire sites were selected across the Pohjanmaa region. Two of the sites were used in Jussila et al. (2024) model training. Reference data were produced for 5x5 kilometer ortophotos, from two different years for some sites. One additional site was visited in the field.

For Flanders, a multitemporal validation set was created, over the same 4 study sites and across 4 different years (2020, 2021, 2023 and 2024). For Slovakia, one site in the Danube floodplain was chosen, and validation polygons were obtained for 2024 during a flood event. In Czechia, two sites were chosen in the Třeboňsko Protected Area. Field points were collected for validation, and orthophotos were used for an additional validation.

For Bolzano, the wetland area in the Lago di Caldaro biotope was chosen as one of the largest and most important wetlands in the province. Validation polygons for the area were defined using a set of provincial aerial photos from 2020. Additionally, parts of the study area were captured using a multispectral drone in August 2025.

For Sweden, three aapa mires were selected and drone images were taken on June 26 2025. For Croatia six Natura2000 sites were selected with temporary ponds assessed in the field (and photographed) in 2022.

Satellite image acquisition

Sentinel-2 L2A images were acquired through the OpenEO platform (Schramm et al. 2021) that is part of the Copernicus Data Space Ecosystem (CDSE). We downloaded single date images that were temporally coincident (or very close) with the reference data (field visit or aerial image acquisition date) and had no clouds over the study site(s).

Creation of reference data

Within each study site, inundated (water) and non-inundated (dry) reference areas were delineated. Reference data were generated using two methods: (1) delineation on high-resolution imagery (e.g., orthophotos or drone data) or (2) direct in-situ assessment of inundation for a point sample.

For some study sites, the delineation of inundated and dry polygons from HR imagery was done in the LabelMe GUI. This Python library allows a user to delineate homogeneous areas in a semi-automatic way using AI-based models and label those areas. We created a Github repository for our implementation of LabelMe (https://github.com/inbo/Biodiversaplus_HabitatPilot-Inundation-SegmentingOrthophotos/tree/master). For Flanders, we created the reference data using semi-automated delineation in LabelMe on orthophotos with a spatial resolution of 15-25cm.

For the Finnish sites, a raster-based validation approach was tested. High-resolution reference rasters (50cm) were produced by unsupervised classification of aerial images stacked with elevation data. The resulting 5-10 clusters were manually labeled as inundated or non-inundated.

For all approaches, to account for areas where interpretation was challenging, three specific ambiguity classes could be added to the reference data:

- Other: This class denotes areas with a land cover not relevant to the target analysis, such as regions obscured by dense tree canopy or artificial infrastructure.
- Uncertain: This class was used for areas that appeared very wet or saturated, but where the presence of actual standing surface water could not be definitively confirmed from the reference imagery.
- Reeds: This class identifies areas covered by reeds or other tall emergent grasses. This vegetation is problematic as it can obscure underlying water, making it invisible to the sensor.

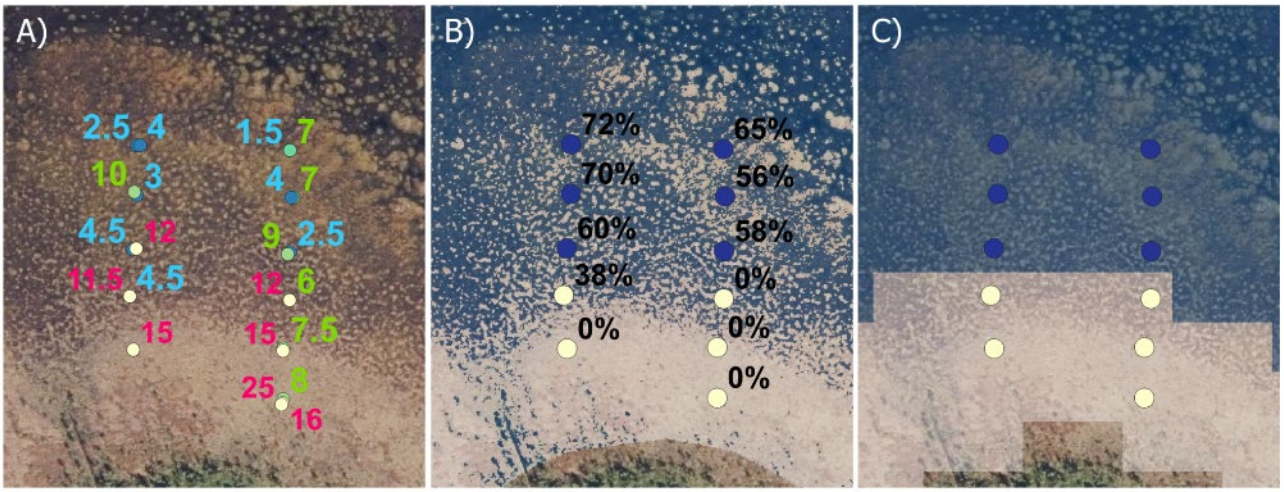
For the Finnish sites, somewhat bolder interpretations were made. Obviously water-saturated dark surfaces (aapa mire flarks, water table depth typically < 5cm) were assigned to the inundated class instead of uncertain, even if clear open water bodies were not present. Habitat mapping information was used to support interpretation.

Field visits in Finland and Czechia were made in August 2025, targeting especially locations that were challenging for aerial image interpretation (vegetated areas, transects along ambiguous wetness gradients). The Finnish field visit site was selected from an area where ortophotos were taken in July during the same summer. Points were labeled as inundated or non-inundated, taking into account the dominant condition within the surrounding 10 meters. Water table depth (WTD) was measured for the point locations in Finland, 2-3 measurements representing the varying water levels

of the heterogeneous microtopography. Dominant surface (inundated or non-inundated) was later calculated based on aerial image HR classification and photos taken at the locations. Point locations dominated by measured open water surfaces (WTD <5cm) were considered inundated.

A2_Fig 1 illustrates the combined use of WTD-measurements and aerial photos.

Field visit in Slovakia was done in September 2024 during a huge flood event. Some 3150 habitats occurring in the area were tracked from the boat using a GNSS device. GPS tracks were imported into NaturaSat software and smoothed in the parts where the original track was reflecting difficulties to access the area. This GPS-tracking approach was tried in the Finnish field site as well, but was abandoned due to high-resolution heterogeneity and ambiguous wetness gradients in mires.



A2_Fig 1 A) measured WTDs (cm) in the field (blue labels for WTD <5 cm), with the aerial image on background, B) aerial image based HR-reference raster and derived percentages of inundated surface in 20m pixel area, C) Jussila model prediction and reference points: dark blue color representing water for reference points and predicted pixel classifications.

The use of drone imagery was assessed to derive precise inundation data for the 3 aapa mires in northern Sweden.We surveyed the central flark areas of three aapa mires in Västerbotten and Norrbotten using a DJI Mavic 3 Multispectral (M3M). Reference water masks were created by setting a site-specific threshold on the NIR band of these drone images.

Data exploration

We generated bar plots showing the total pixel count for each reference class to inspect the class imbalance in the reference data and to assess the prevalence of uncertain interpretations. The prevalence and effects of the mixed pixel problem were assessed by analysing the reference polygon delineations and the HR reference rasters in Sentinel-2 pixel 10-20m grids.

Using the pixel-based validation approach,we calculated the dominant label (majority voting) as well as the purity class (> 90% = pure; 60-90% = mixed; <60% = very mixed) for each Sentinel-2 pixel.

We also created boxplots for the Sentinel-2 reflectance bands and derived indices. This visualization allowed us to visually inspect the spectral separability between the defined reference classes (e.g.,

water, land, reeds) and identify potential confusion between them. Also, the usability of water index thresholding for wetlands could be assessed.

Model validation

Two inundation models were tested: Water in Wetlands (WiW) and Jussila's model.

The Water in wetland (WiW) model was trained in Camargue wetlands in France (Lefebvre et al. 2019). This very simple model consists of a single decision tree that uses only two optical bands, a Near Infrared (NIR) band and a Short Wave Infrared (SWIR) band.

In that model, a pixel is considered as inundated if the NIR reflectance ≤ 0.1804 and SWIR reflectance ≤ 0.1131. The authors reported a good overall accuracy of 94.1%. For Sentinel-2, band 8 serves as the NIR band and band 12 as the SWIR band.

Jussila's model (from Jussila et al. 2024) was trained in boreal aapa mires in Finland, following the same method as Lefebvre et al (2019). The model, a single decision tree, relies largely on the NIR (band 8), SWIR (band 11 & band 12) and red (band 4) bands but it is slightly more complex as the WiW model and also contains the Modified Normalized Difference Water Index with band 12 (MNDWI12) index.

MNDWI12 is calculated as

$$\frac{green - SWIR2}{green + SWIR2}$$

with band 3 as the green band and band 12 as the SWIR2 band for Sentinel-2.

This model can detect both open water, and surfaces in which the water table is close to the mire surface level (0-5 cm depth). The authors reported an overall accuracy of 93.6 % for Sentinel-2.

Both models were applied on Sentinel-imagery downloaded locally for the individual study sites (see Table 1) across different regions in Europe. The models were applied in R for validation, but we also tested their application on a larger scale through their integration into Google Earth Engine. The success of this test confirmed the feasibility of upscaling these simple models to the national and potentially even European level.

For generating validation samples, we used a pixel-based approach. It calculates the dominant label within each Sentinel-2 pixel as the reference class that occupies the largest proportion of the pixel. The latter approach allows for the calculation of a purity level (i.e. the percentage of the pixel taken up by the dominant label), which was done for some study sites.

For validation, uncertain and mixed pixels were removed from reference data prior to Macro-F1 calculation.

Special use case: Superresolution

As a special use case, we tested the performance of the Jussila model on super-resolution Sentinel-2 images generated by the Sentinel-2 Deep Resolution 3.0 model, which is a model similar to the Enhanced Super-Resolution Generative Adversarial Network (ERSGAN). For more information about this model, see the white paper of Akhtman (2023).

We generated super-resolution images for the 4 study sites of Flanders that temporally coincide with the original Sentinel-2 images used for the validation of the inundation models. We then applied the Jussila model on these superresolution images and checked the impact on the spectral model performance and on the spatial delineation of water patches and on the overall inundation estimates.

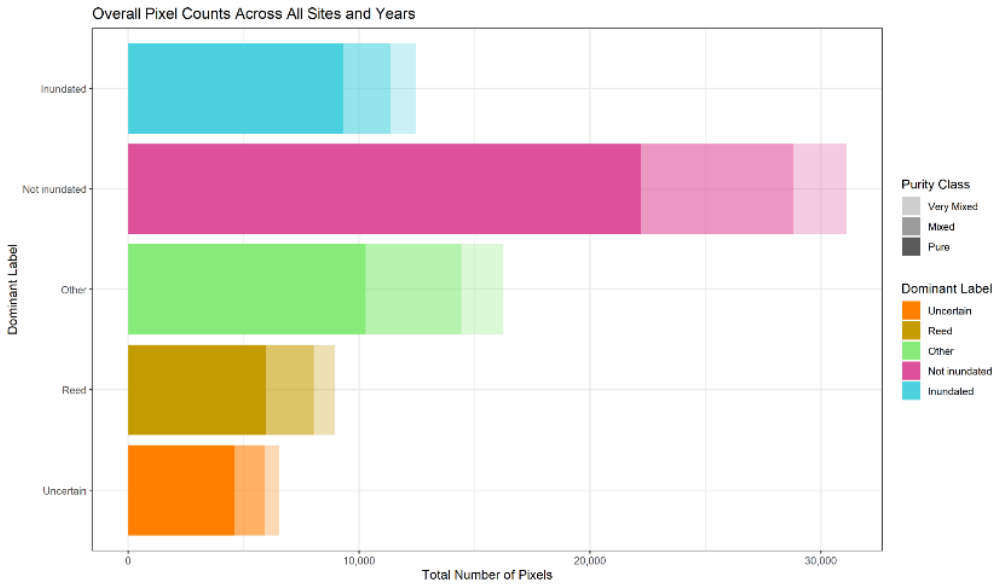
For each combination of study site and year (14 in total), we calculated the difference in performance (precision, recall, F1-score) between the super-resolution and the original Sentinel-2 images as a means of assessing the spectral performance. To assess the spatial performance of the super-resolution versus the original, we used the differences in IoU and Dice index.

Finally, we also calculated the percentage of the area that was classified as inundated for the super-resolution and the original Sentinel-2 images. We plotted this against the reference percentage and calculated the R^2 relative to the 1-1 line. All performance indicators were calculated on the pixels that had Inundated or Not inundated as their dominant label in the reference dataset. Pixels with other reference labels were disregarded.

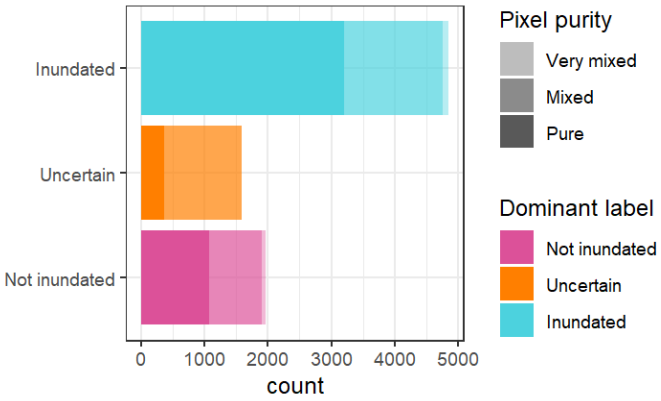
Results

Data exploration

The bar plots in A2_Fig 2 and A2_Fig 3 shows the distribution of the dominant labels and the purity classes in Flanders and Finland. We are faced with a somewhat imbalanced dataset, with more areas that are not inundated than that are inundated in Flanders, and vice versa in Finland. Most pixels are pure, but there is a significant proportion of mixed pixels (37% in Finland) across all dominant labels, of which a small minority are very mixed. In the Finnish reference data, “Uncertain” was the dominant label in 19 % of pixels, for Flanders this was 8.7%. Combined, 49 % of pixels were either mixed or had uncertainty in labeling in the Finnish reference data compared to 37% in Flanders (which has more classes).

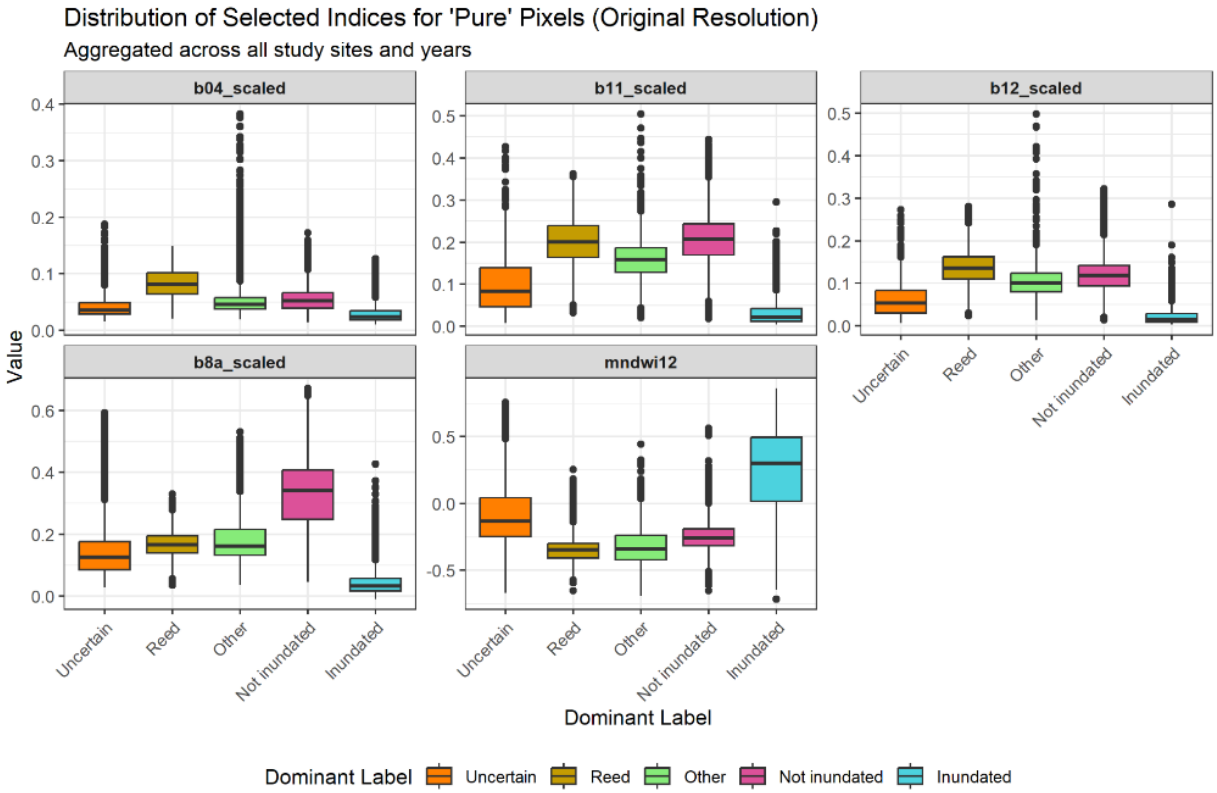


A2_Fig 2 Barplot of the distribution of the dominant labels and the purity classes in Flanders.

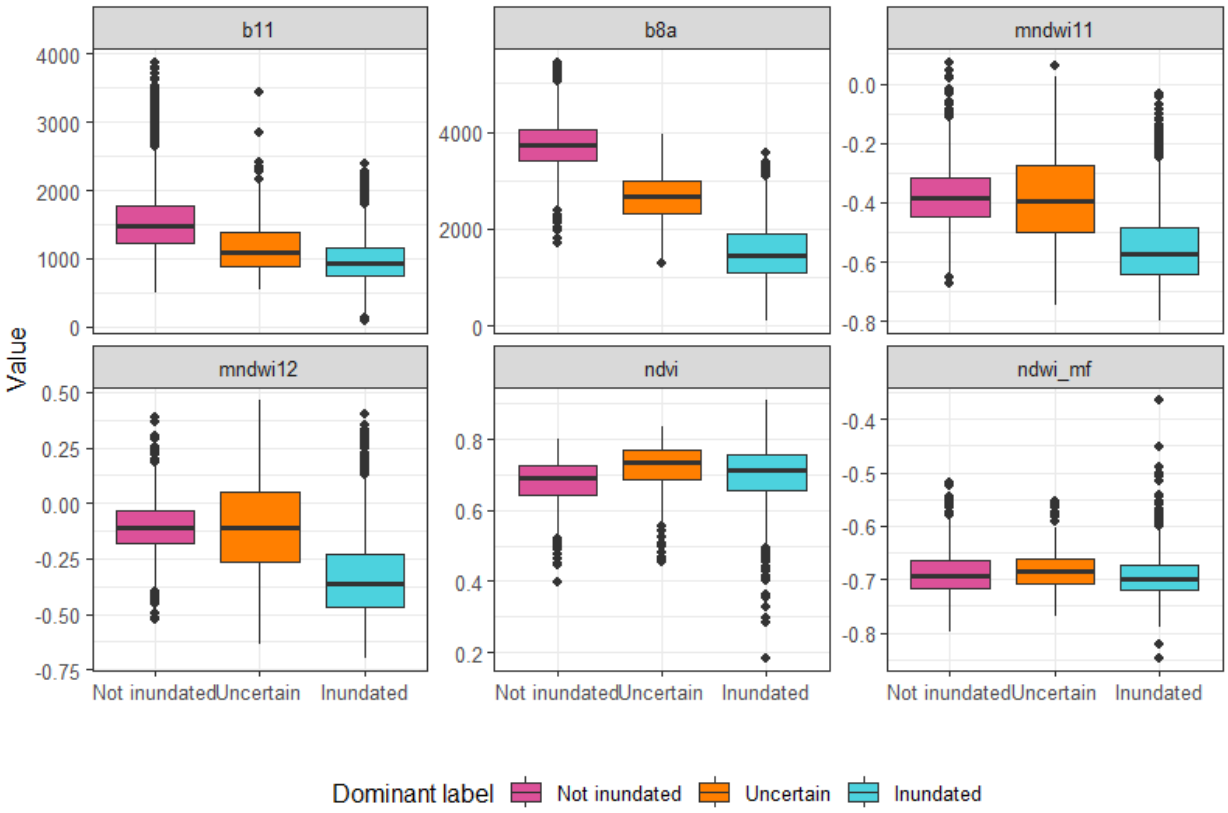


A2_Fig 3 Barplot of the distribution of the dominant labels and the purity classes in Flanders.

A2_Fig 4 shows the boxplots for Flanders of the different Sentinel-2 bands and derived indices used in the Jussila or WiW inundation models, for the 5 dominant labels. We see that the Inundated and the Not inundated pixels are well separable across many bands, and that the Uncertain label is an intermediary between these two extremes in the bands used in the classification models (band 8a, band 11, band 12). The MNDWI12 index shows a similar distribution. The labels Reeds and Other seem to be more similar to Not inundated, especially for band 4, band 12 and the index.



A2_Fig 4 Boxplots of all reference labels across the different bands used in the inundation models for all study sites and years in Flanders combined.



A2_Fig 5 Boxplots of the different bands used in the inundation models across all study sites in Finland.

A2_Fig 5 above shows exploration of band values and water indices for the validation data from Finnish aapa mires (including pure pixels only, threshold 90%). Non-inundated and inundated pixels are separable with bands b11 and b8a and uncertain reference points fall in between. The modified water indices (mNDWI) show separable distributions as well. Uncertain pixels have a similar distribution to non-inundated pixels. Unexpectedly, instead of high values, the data shows low negative mNDWI values for inundation in aapa mires. McFeeters water index (NDWI) and vegetation index (NDVI) show a similar distribution of values across classes, with poor separability.

Model validation

A2_Table 2 summarizes the macro-F1 scores for each of the regions involved in this subtask. Classification accuracy has similar levels for both models, with scores 0.74-0.92 for Jussila and 0.64-0.93 for WiW model. However, certain differences between models were observed. For Czechia, Bolzano and Finland, the Jussila model seemed to be more sensitive to water, classifying generally more pixels as water (Figure 6) and detecting water in mixed pixels (if water covers more than half of the pixel). For Flanders, the WiW model seemed more water-sensitive, predicting more water pixels. Also for the class “uncertain”, Jussila model predicted more water.

A2_Table 2 Overview of model validation results (macro-F1 scores) across the different regions involved in the inundation subtask.

Region	Validation approach	N (points)	Macro F1 Jussila	Macro F1 WiW
Finland	Aerial images	4285	0.8213	0.8139
Finland	Field validation	30	0.917	0
Flanders	Aerial images	75317*	0.8917	0.8871
Czechia	Field validation	279	0.7408	0.6983
Slovakia	Field validation	230	0.9128	0.9251
Bolzano	Aerial images	6673	0.7665	0.6425
Croatia	Combination	167	NA**	NA**

*For Flanders, reference data were gathered for 4 years in the same areas, leading to a higher N.

** For Croatia, there was no overlap between the reference and predicted water areas.

Model performance results vary between the aerial image validations and field validations. For Finland, Jussila model performance was higher with field validation, whereas WiW model was unable to detect any watery areas in the field data points. For Czechia and Bolzano, the model performances for field reference were relatively low compared to other cases. For Czechia, this was probably due to the focused collection of reference data in problematic areas. In the case of Bolzano, this was most likely due to the low total number of inundated pixels in the wetland.

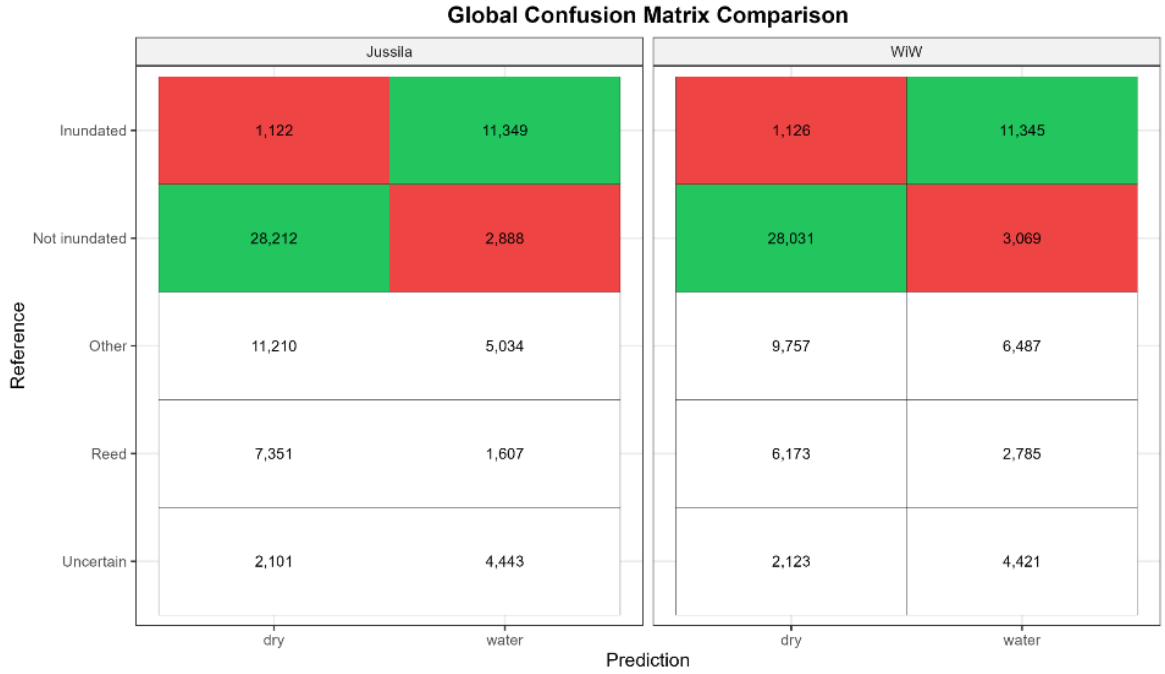
For Croatia, the predicted water pixels do not overlap with the pixels labeled as water in the reference dataset. This is probably due to the small size of the habitats mapped, which does not reach the minimum mapping unit of 1 Sentinel-2 pixel (10 x 10 meters), frequent presence of individual smaller trees around the inundated areas, combined with some uncertainty about the exact location of the inundated areas, which were based on non-georeferenced field assessments.

Tree-cover was observed as a limitation for water detection especially for the Jussila model, as was reported in Jussila et al. (2024). Tree-covered areas, particularly coniferous trees, were systematically misclassified as water. WiW model did not seem to have the same bias, but based on the reference data it is unsure if presence of water would be detected under trees. Also tree-shadows were misclassified as water. For the accuracy analysis, tree-covered pixels were masked out of the data. Accuracy metrics reflect performance on open areas.

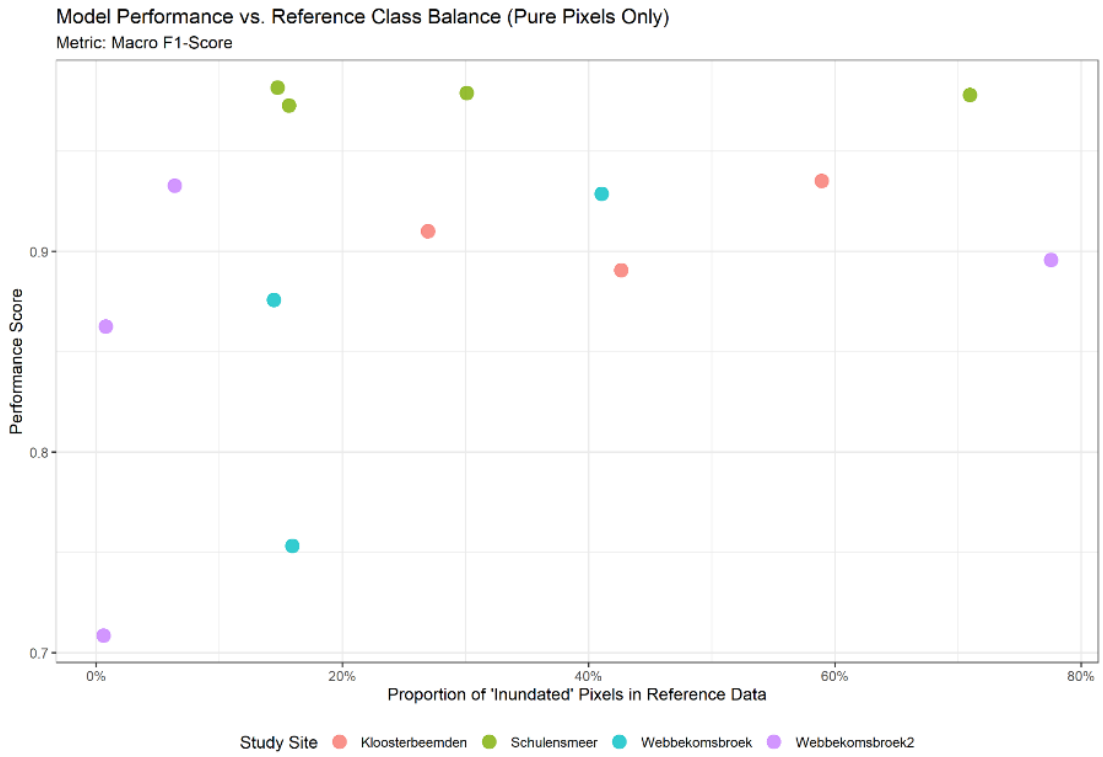
Also other high vegetation affected water detection. In Czechia, and Slovakia, the WiW model did not detect water under high reedbeds. For Flanders, some water was predicted under reedbeds by both models (see Fig.6). However, our ability to validate the performance on reedbeds was limited, due to difficulty in producing interpretations for reference data in these areas.

For Flanders, it seems that the WiW model is more sensitive to water, as it classifies more pixels as water across all reference labels (A2_Fig 6), except for the Inundated and Uncertain labels. At the

study site level, the performance (F1-score) of both models (results only shown for the Jussila model in A2_Fig 7) seems to increase with higher inundation levels.



A2_Fig 6 Confusion matrix of the inundation classification models, for all validation pixels in Flanders (across all study sites). Wrong classification is indicated in red, correct classification in green. The classes in white are not directly linked to either of the model output classes and could not be 'validated'.

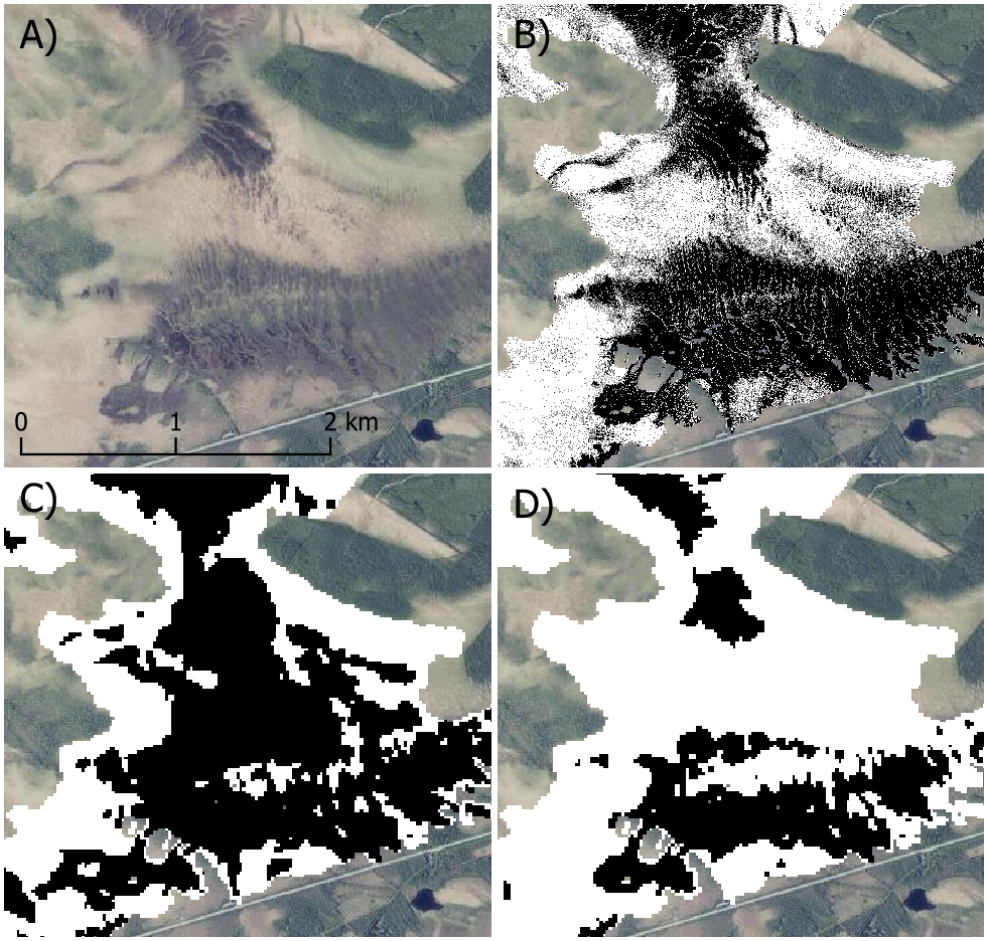


A2_Fig 7 Plot showing the macro-F1 score as a function of the proportion (of the study site) inundated at the time of validation, for all study sites across the validation years in Flanders.

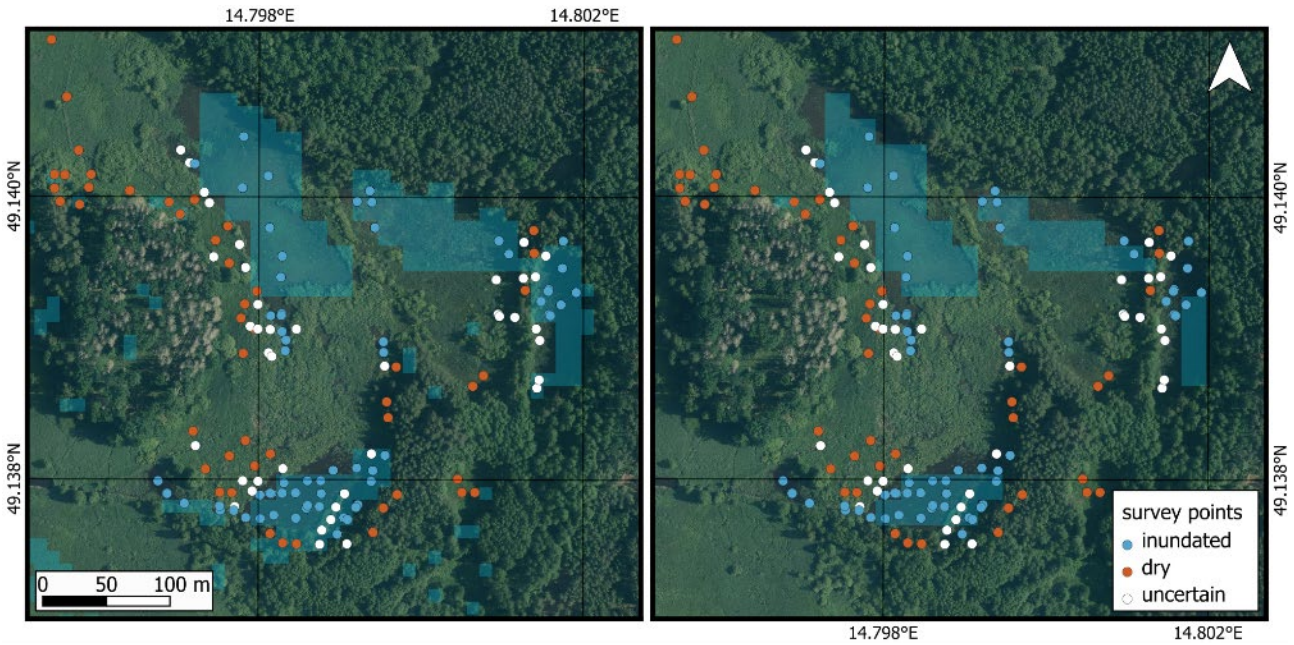
Some visualizations of the model outputs are consulted below; for the Hirvisuo study site in Finland (A2_Fig 8), the Kramářka study site in Czechia (A2_Fig 9), the Bodíky study site in Slovakia (A2_Fig 10) and the Lago di Caldaro wetland area in Bolzano (A2_Fig 11).

For Sweden, the comparison of the water cover detected by the Sentinel-2-based Jussila inundation index and by the drone-based approach shows good spatial agreement (A2_Fig 11). The wettest areas largely coincide, with discrepancies - most evident at Jörn 02- likely caused by the temporal mismatch. This is because obtaining a cloud-free Sentinel-2 scene near the drone acquisition date was difficult.

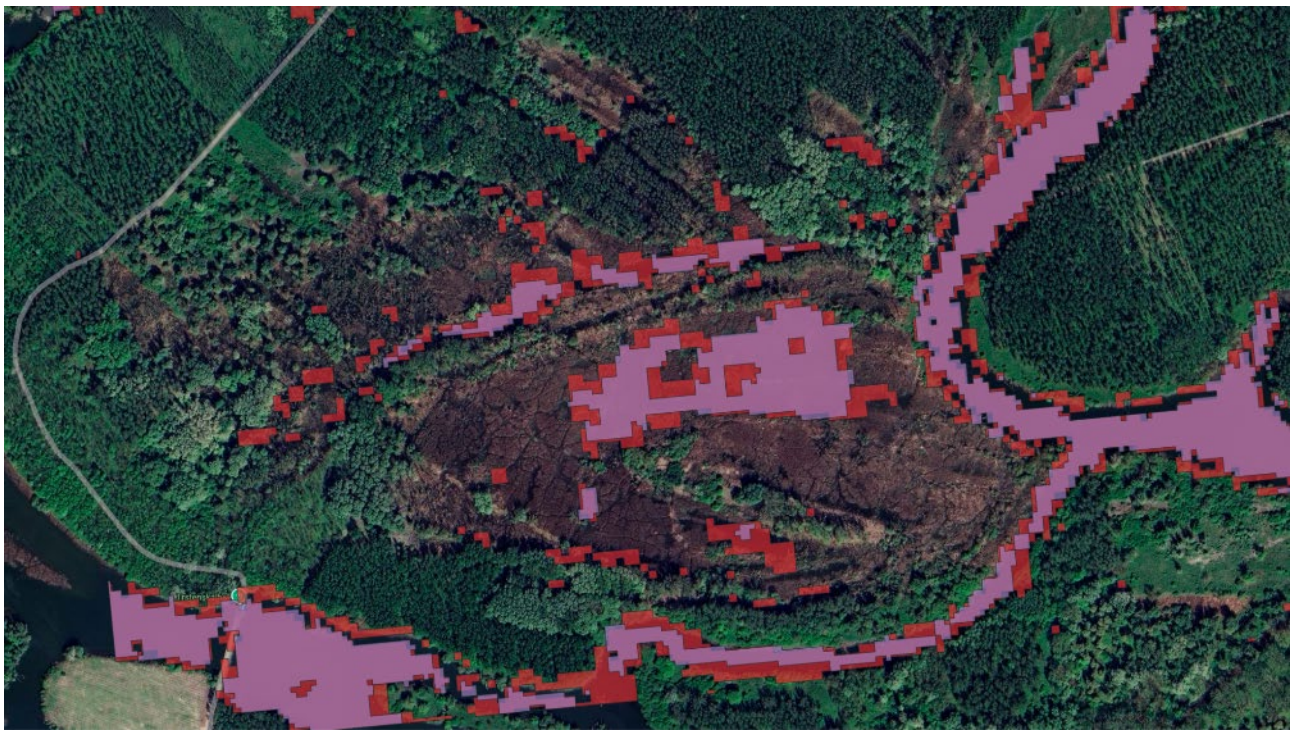
Appendix 2. Inundation mapping



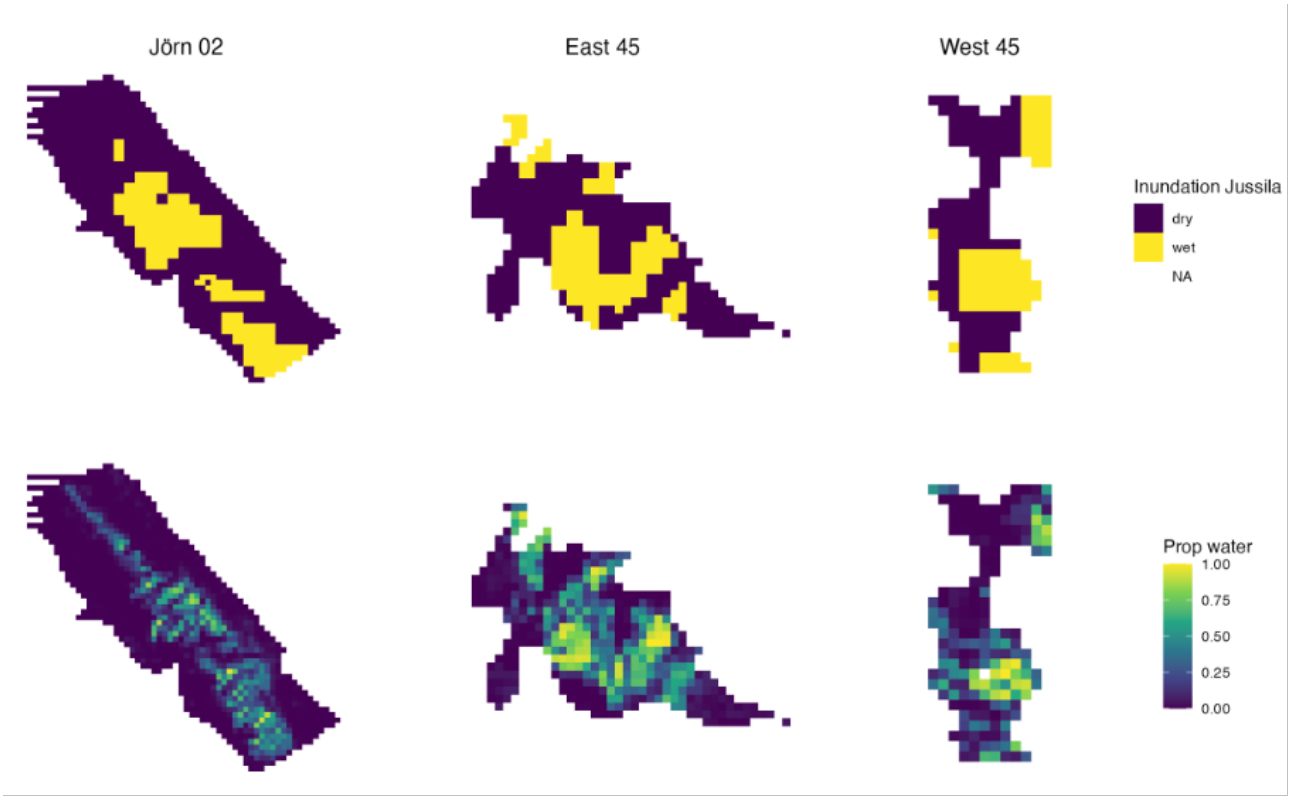
A2_Fig 8 A) aerial image of Hirvisuo aapa mire site in Finland, B) High-resolution inundation reference raster based on aerial image, C) Jussila model inundation prediction, D) Lefebvre WiW model prediction. Jussila model seems to overpredict inundation, whereas WiW underestimates inundated area



A2_Fig 9 Jussila model output (left) vs. Lefebvre model output (right) for the Kramářka study site in Czechia (inundated = blue).



A2_Fig 10 Jussila model output (inundated = red) vs. Lefebvre model output (inundated = pink) for the Bodíky study site in Slovakia, showing a larger area identified as inundated by the Jussila model.



A2_Fig 11 Visual comparison of the water delineation between the Jussila model (top) and the drone-based assessment (bottom) in three aapa mires in Sweden

Use case Super-resolution

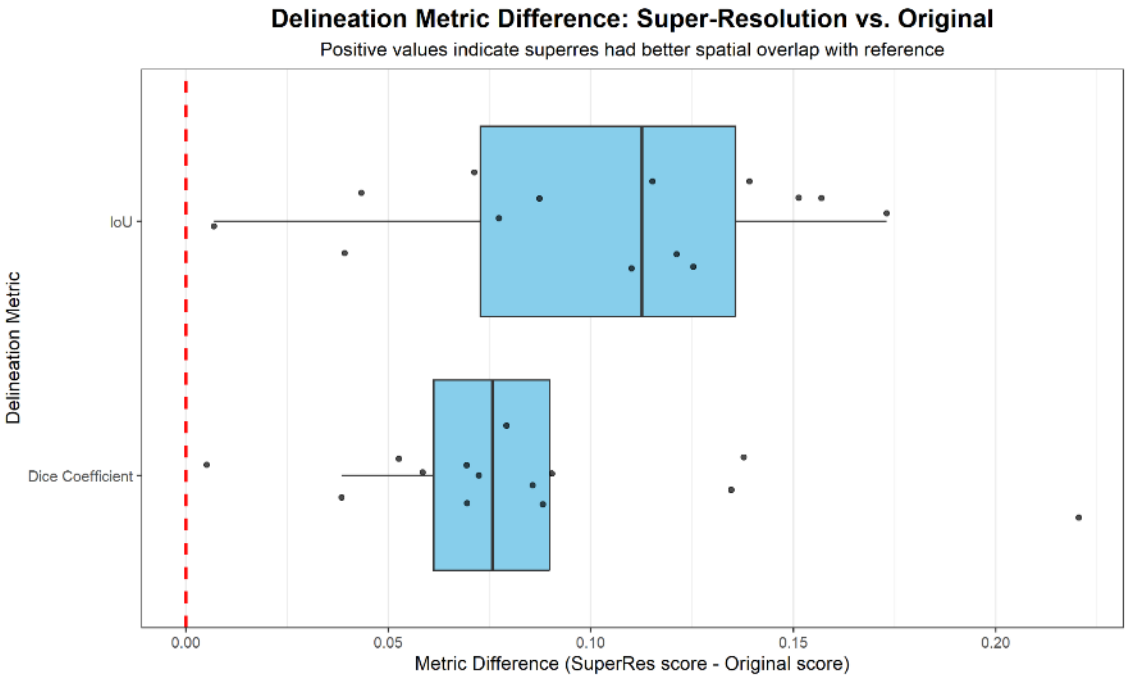
The spectral (pixel-level) model performance of the super-resolution images shows no significant differences with that of the original Sentinel-2 images (data not shown), for none of the indicators. This result underscores the good spectral performance of the super-resolution images, as they seem to introduce no spectral deviations.

The spatial model performance on the other hand is significantly better for the super-resolution than for the original images (see A2_Fig 12). The delineation of super-resolution-based water bodies thus has a better spatial correspondence to the delineation of the reference water bodies. The original images have a slightly higher R^2 which means that their estimates correspond slightly better to the reference percentages (see A2_Fig 13).

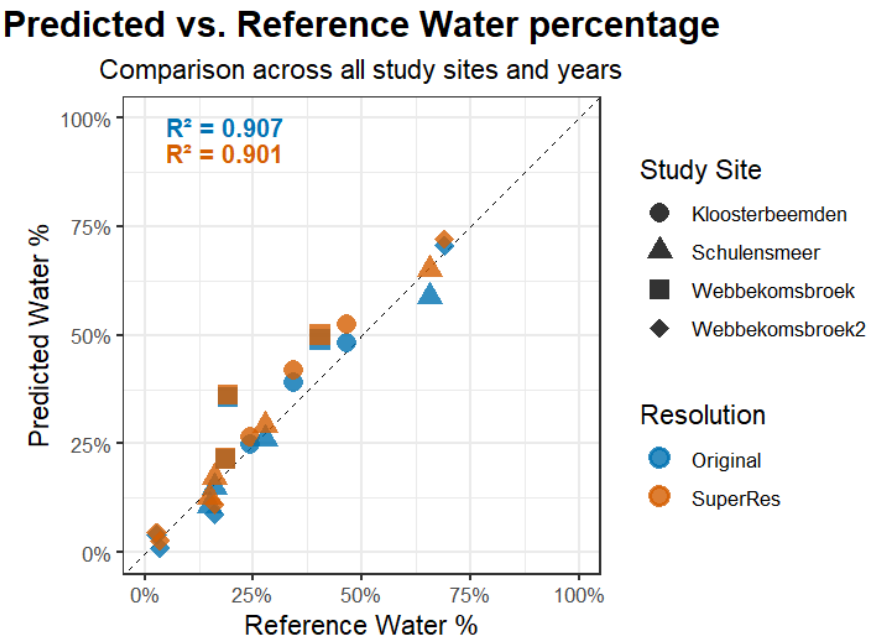
A2_Fig 14 shows some examples of classification results for some zoom-ins in our study sites. They show a better delineation of the actual water bodies, especially the smaller or narrower ones. But they also show some scattered 'hallucinated' water pixels that seem to be created by the upscaling of the images.

Similar results were found for the study site in Bolzano where a section of the inundated area were narrow water tracks that could not be discerned and classified as inundated using the resolution of the original Sentinel-2 images. The improved spatial resolution of the super resolution image led to more of these pixels being un-mixed and to some of the water tracks being classified as inundated (A2_Fig 15).

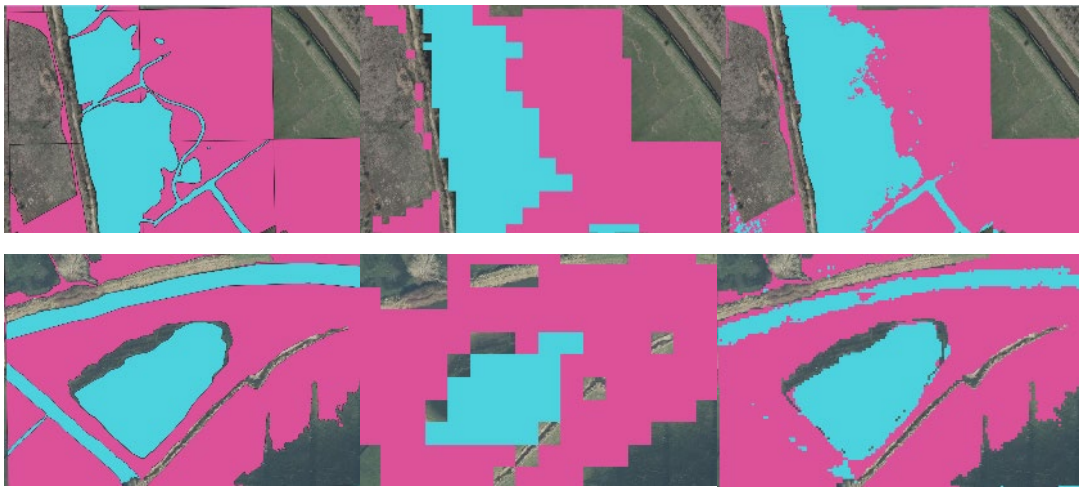
In Finland, the super resolution testing was done in a part of Hirvisuo aapa mire, where inundated flarks and drier areas form very fine resolution surface patterning (A2_Fig 16). Due to generally moist mire conditions, inundation in mixed pixels is detected, even in the original resolution Sentinel-images. However, drier strings blend in the inundation-dominated mixed pixels and the total area of inundation is overestimated. Superresolution classification shows more fine resolution heterogeneity, ending up with more accurate prediction of the total inundated area compared to the prediction with the original resolution.



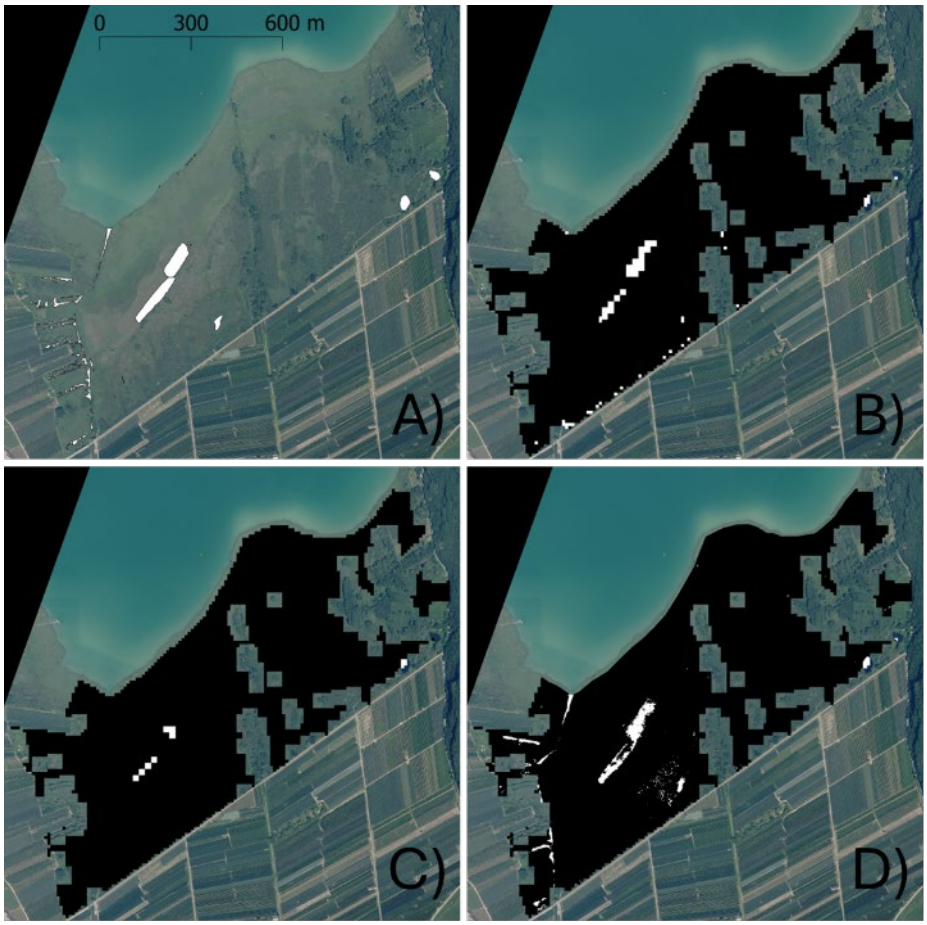
A2_Fig 12 Differences in per-pixel performance of the Jussila model between the superresolution and the original resolution, for all study sites and years in Flanders.



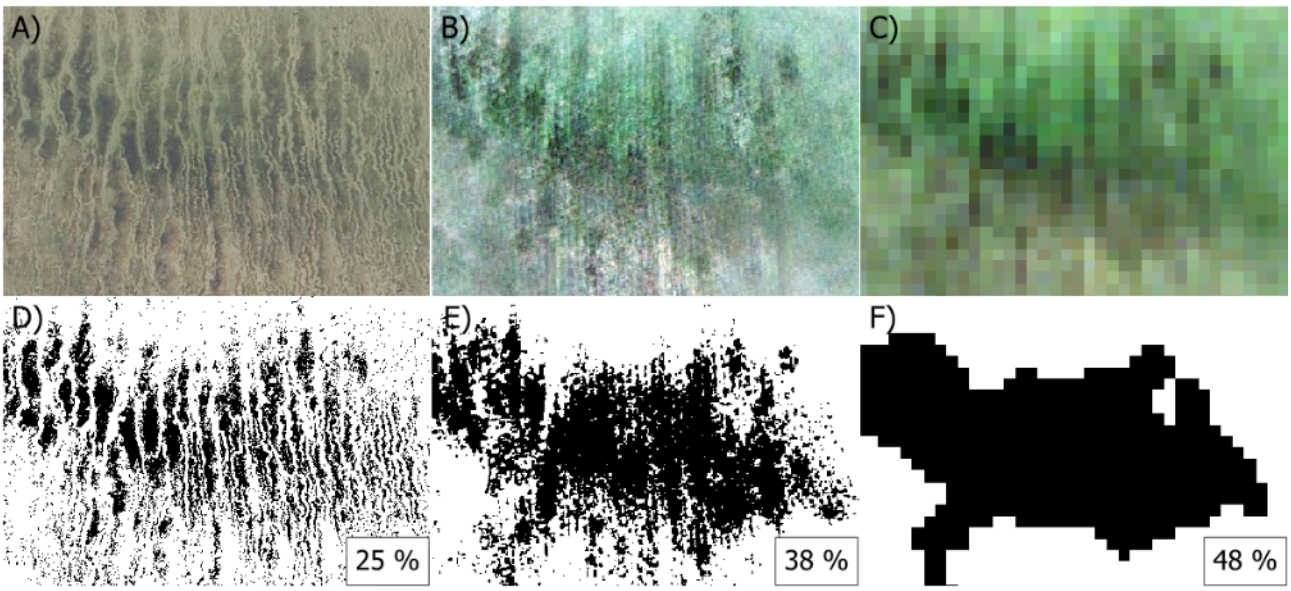
A2_Fig 13 Scatter plot of the reference area water cover (%) versus the predicted water area (%) by the Jussila model for the original resolution and the superresolution for all study sites and years in Flanders. The indicated R² is relative to the 1-1 line.



A2_Fig 14 Reference labels (left), Sentinel-2 classification (middle) and super-resolution classification (right) for study site Webbekoms Broek in 2020 (top) and study site Schulensmeer in 2021 (bottom).



A2_Fig 15 A) Aerial image of Lago di Caldaro wetland area in Bolzano with inundated areas segmented and visualized in white, B) Jussila model inundation prediction, C) Lefebvre WiW model prediction. D) Jussila model inundation prediction on super resolution image. Areas with high vegetation, including a surrounding buffer area, have been masked out.



A2_Fig 16 A) aerial image of a part of Hirvisuo wetland in Finland, B) Super resolution rgb image, C) original Sentinel rgb image, D) reference classification of wet areas (black) based on aerial image unsupervised classification and visual interpretation, E) Jussila model inundation prediction on super resolution image, F) Jussila model inundation prediction on original Sentinel-2 image. Numbers in boxes present the calculated percentage of inundated area in the image area.

Conclusions

Both classification models (Jussila's model and WiW model) had a good performance across most of the validated regions. The approach proved effective for capturing the general extent of significant flooding events. Our analysis showed that model performance (Macro F1 score) was a function of the inundation level, with scores generally increasing in more heavily inundated areas. This suggests that as flooding becomes more widespread and homogeneous, detection improves, an increase seemingly driven by the higher prevalence of pure, open-water pixels.

However, the validation also highlighted some clear challenges and limitations.

Spectral confusion is the first challenge. Tree shadows were frequently misclassified as water, leading to false positives. Conversely, the presence of reeds or other tall, emergent grasses effectively obscured underlying water, leading to false negatives and an underestimation of the true flood extent. Saturated grasslands (or 'drasland'), where the water table is high but not necessarily forming standing surface water, also proved to be a consistent source of interpretation error.

Besides spectral confusion, spatial resolution was another limitation. The 10-meter pixel size of Sentinel-2 is often too coarse to detect small or narrow inundated areas, such as flooded ditches, small streams, river branches or the fragmented edges of a larger flood. While superresolution techniques can improve the visual delineation of small water bodies, they also introduced significant artifacts, such as 'hallucinated inundations' in completely dry areas. The integration of a detailed Digital Terrain Model (DTM) could improve the results, helping to filter out implausible inundated areas (e.g., on steep slopes) and refine flood boundaries.

Ultimately, this study underscores a fundamental uncertainty in all flood delineation, which was evident not only in the satellite data but also on high-resolution orthophotos and even during in-situ field checks.

Discussion

In this subtask, we assessed the accuracy and transferability of two models for classifying inundated areas. We found that these models (Jussila's and WiW), despite being trained on a single ecosystem, showed strong potential for EU-wide inundation monitoring across a range of open ecosystems.

We selected these models due to their ease of implementation and specific design for delineating dynamic wetland inundation - a central focus of the Habitats Pilot. While more advanced machine learning or deep learning models might improve performance, they typically require large training datasets and are often more difficult to transfer to new areas (Vali et al., 2020).

Despite the models' promise, moving towards an operational implementation requires addressing the primary limitation of optical Sentinel-2 data: cloud coverage. We recommend that future research incorporate radar satellites like Sentinel-1, which was outside the scope of this project.

Beyond this, our pilot identified several core challenges in the validation process itself, discussed here below.

The inundation gradient vs. binary classification

A key conceptual challenge is that inundation is often a continuous spatial gradient, not a discrete "water/no-water" boundary. This ambiguity directly conflicts with a binary classification approach. This raises two critical points:

Accuracy metrics: The ambiguity calls into question the use of binary accuracy metrics. Given the fuzzy nature of real-world inundation, misclassifications near a class boundary are not necessarily a serious model failure.

Monitoring approach: Future monitoring might be more robust if it shifted from mapping horizontal extent to monitoring vertical water level, though the feasibility of this with optical RS data remains uncertain.

The subjectivity of reference data

Validation based on aerial imagery is cost-effective and scalable, but it suffers from interpretation uncertainty, which decreases the reliability of accuracy assessments. This subjectivity in labeling creates significant issues:

Interpreter bias: Optimistic versus pessimistic interpretations led to large differences in estimated water extent (and thus, accuracy metrics) in the Finnish reference data.

Harmonization: We also noted different interpretation "attitudes" between partners—such as how sensitively dark, non-open-water surfaces were labeled. Harmonizing this subjective interpretation, which may be linked to different habitat types, is a major challenge.

Ambiguous surfaces: Open water bodies were easy to delineate, but ambiguous, dark, or watery-appearing surfaces were not. These areas—which include thin water layers, mudflats, or water-

saturated peat and moss—are ecologically inundated but extremely difficult to label correctly without a field visit.

Limited multi-temporal validation: Beyond spatial performance, a model's suitability for environmental monitoring depends on its ability to accurately reflect changes in the water regime, both seasonally and inter-annually. Our capacity for such multitemporal validation in this pilot was limited.

Recommendations for future monitoring

Based on the challenges listed above, we offer two primary recommendations:

Explore a three-class model or continuous water level monitoring: Given the difficulties with binary detection, a potential solution is to adopt a model with three classes: 1) open inundated water, 2) semi-inundated/water-saturated, and 3) not inundated. Alternatively, we could attempt to model groundwater levels in a continuous way, but this would require a different type of input data, preferably radar data (e.g. Sentinel-1). As this question comes with its own set of challenges, we excluded it from this sub-task.

Adopt a hybrid validation strategy: To address data uncertainty, we strongly recommend combining drone imagery with targeted, simultaneous field-verified points. This approach marries the objectivity of field data with the upscaling potential of drone imagery, enabling the creation of larger, more robust validation datasets. If possible, all field data collection should be timed to coincide with aerial or satellite image acquisition.

Use/validate existing datasets: There are some large-scale inundation/water products at the European or global level. A first example is the Global Surface Water Explorer (GSWE) product suite (Pekel et al., 2016). It is a high-resolution (30m) global dataset developed by the European Commission's Joint Research Centre and Google that maps the location and temporal distribution of surface water. Derived from the entire Landsat archive (1984–present), it provides key data layers on water occurrence, seasonality, and long-term changes.

A new pan-European HRL 'Water Cover Duration' is a planned dataset that will provide a yearly aggregated layer showing the duration of water cover. This product is derived from the pan-European Water and Ice Cover product and is scheduled for its first release (covering 2024/2025) in January 2026 (Copernicus, 2025). The pan-European Water and Ice Cover is a near-real-time (NRT) product that provides daily, pixel-based information on water and ice presence in rivers and lakes. This new product, based on both Sentinel-1 (radar) and Sentinel-2 (optical) data at 20m/60m resolution, replaces the older "River and Lake Ice Extent" product and has been in production since mid-January 2025.

These datasets require validation to confirm they accurately capture dynamic wetland inundation, and not just the more static water found in permanent water bodies.

Explore super-resolution further: With respect to super-resolution, we recommend further research on its added value for the delineation of flooded areas (in wetlands). For this purpose, we can either use upsampled Sentinel images or upsampled PlanetScope images. Planet SuperRes is an AI-enhanced product line that provides a higher-resolution version of its daily PlanetScope imagery. It uses a proprietary artificial intelligence model, trained by comparing its standard 3.7m PlanetScope data with its much higher-resolution 50cm SkySat imagery. The AI model effectively

"up-samples" the daily PlanetScope imagery, resulting in a new, derived dataset with a significantly sharper 1.2m resolution. This allows users to get near-daily imagery with much greater detail than standard PlanetScope data.

Further evaluation of predictive epistemic uncertainty:

Both the Jussila model and the WiW are trained and evaluated on a limited training, validation and test dataset. Extrapolation of these models to other time periods and spatial regions creates additional epistemic uncertainty. This is especially challenging in tree-based models (Ellis, 2016). Machine learning models that provide prediction probabilities can be used to calculate the total uncertainty (Shannon Entropy (H)) of the predictions, which indirectly can be used to quantify epistemic uncertainty. Shannon entropy is a good estimate of the average total uncertainty, which is equal to the sum of the aleatoric and the epistemic uncertainty.

Another way to evaluate epistemic uncertainty is by comparing the multivariate distribution of the model's training data in feature space, and compare it with new data (from other regions or time periods) to detect out-of-sample/out-of-distribution pixels. We recommend further research in trying to quantify these uncertainties when upscaling these models to EU-wide inundation monitoring.

Supplementary materials

Sample code supporting this subtask is publicly available in the Biodiversa Habitat Pilot repository (INBO, 2025).

References

Akhtman, Y.. 'Sentinel-2 Deep Resolution 3.0' (2023). Medium . https://medium.com/@ya_71389/sentinel-2-deep-resolution-3-0-c71a601a2253

Copernicus (2025). https://land.copernicus.eu/en/products/water-bodies?tab=roadmap

Ellis, P. 'Extrapolation Is Tough for Trees!' (2016) Free Range Statistics. https://freerangestats.info/blog/2016/12/10/extrapolation.html

INBO. (2024). BiodiversaHabitatPilot (GitHub Repository). Available at: https://github.com/inbo/BiodiversaHabitatPilot

Jussila, T., Heikkinen, R.K. Anttila, S. e.a. (2024). Quantifying Wetness Variability in Aapa Mires with Sentinel-2: Towards Improved Monitoring of an EU Priority Habitat. Remote Sensing in Ecology and Conservation 10, 172-87. https://doi.org/10.1002/rse2.363

Lefebvre, G., Davranche, A. Willm, L. e.a. (2019). Introducing WIW for Detecting the Presence of Water inWetlands with Landsat and Sentinel Satellites. Remote Sensing 11, 2210. https://doi.org/10.3390/rs11192210

Pekel, JF., Cottam, A., Gorelick, N. et al. (2016) High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422 . https://doi.org/10.1038/nature20584

Schramm, M., Pebesma, E., Milenković, M., Foresta, L., Dries, J., Jacob, A., ... & Reiche, J. (2021). The openeo api–harmonising the use of earth observation cloud services using virtual data cube functionalities. Remote Sensing, 13, 1125. https://doi.org/10.3390/rs13061125

Appendix 2. Inundation mapping

Vali, A., Comai, S., & Matteucci, M. (2020). Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review. Remote Sensing, , 2495. https://doi.org/10.3390/rs12152495

Appendix 3. EU Grassland watch validation

Full title: Testing harmonized validation approaches on interim EU Grassland Watch Typology products

Contributors:

- Oosterlynck Patrik patrik.oosterlynck@inbo.be (Research Institute for Nature and Forest - INBO)
- Verbesselt Sebastiaan sebastiaan.verbesselt@inbo.be (Research Institute for Nature and Forest - INBO)
- Albin Bjärhall albin.bjaerhall@eurac.edu (Eurac Research)
- Jesper Erenskjold Moeslund jesper@ecos.au.dk (Aarhus University)
- Hans Gjardfell hans.gardfjell@slu.se (Swedish University of Agricultural Sciences)
- Andrej Halabuk andrej.halabuk@gmail.com (Slovak Academy of Sciences)
- Vít Ježek vit.jezek@aopk.gov.cz (Nature Conservation Agency of the Czech Republic)
- Martin Kelko martin2kelko@gmail.com (Slovak Academy of Sciences)
- Tamara Kirin tamara.kirin@mzozt.hr (Institute for Environment and Nature, Ministry of Environmental Protection and Green Transition in Croatia)
- Dan Leština dan.lestina@aopk.gov.cz (Nature Conservation Agency of the Czech Republic)
- Jakub Rataj jakub.rataj@aopk.gov.cz (Nature Conservation Agency of the Czech Republic)
- Tomás Rusnak ruso.tomas@gmail.com (Slovak Academy of Sciences)
- Maria Sibikova maria.sibikova@savba.sk (Slovak Academy of Sciences)
- Jozef Sibik jozef.sibik@savba.sk (Slovak Academy of Sciences)
- Jana Spurelova jana.spulerova@savba.sk (Slovak Academy of Sciences)

Abstract

The aim of this subtask in the Habitat Pilot was twofold: firstly, to compare the accuracy of the interim EU Grassland Watch (EUGW) Typology and Land Cover component across various European study areas, and secondly, to test complementary validation techniques. Three standard validation approaches were applied. Additionally, explorative analysis of the temporal trajectories in the yearly composites for the 2016-2023 era was also undertaken. Firstly, indirect validation using ortho-imagery confirmed high overall grassland detection accuracy, yielding an average F1 value of 0.88 across 13 sites, although validator bias and the ambiguity of complex class definitions (e.g. sparsely wooded grassland and forest fringes) affect interpretation of the validation results. Secondly, using local reference datasets proved essential for identifying more detailed recurring misclassifications, despite limitations related to ad hoc availability and the temporal mismatch of these kinds of local datasets. Thirdly, in situ fieldwork revealed a high variance in accuracy for specific EUNIS level 2 classification throughout the tested areas. Applying these complementary validation methods on the local scale revealed several aspects where model output can be improved. For example, from the tested sites, we argue that lower model complexity and higher accuracy can potentially be reached by unambiguously (re)defining certain problematic classes (e.g. wooded pastures) or even narrowing down the scope of the targeted grassland types (e.g. EUNIS level 2 class forest clearings and

fringes). Overall, while the EUGW Typology is successful in general grassland detection, the current standalone product lacks the differentiation between relatively low-productive, species-rich grasslands and more intensively used, species-poor, agricultural grasslands. Although in line with EUNIS class definitions, this is a critical drawback for most national and EU-level grassland biodiversity monitoring requirements. In order to further maximize the added value of EUGW products towards these needs, we believe the intended integration with the EUGW Management and Productivity components to be a promising approach.

Acknowledgement

We are very grateful to the EUGW project team for their crucial support throughout this project. We specifically acknowledge Karl Ruf, Jan Misurec and Stefan Kleeschulte for providing timely access to interim data blocks and taking into account our selection of monitoring sites data production process. Furthermore, we thank EUGW for their valuable assistance in understanding their data products, for delivering comprehensive technical documentation and Heinz Gallaun from Johanneum Institute for presenting the project's internal validation workflow. We would also like to thank Jean-Marc Couvreur from the Department for the Study of the Environment and Natural Assets (DEMNA) of the Walloon region in Belgium for sharing a reference data set used in this work.

Context and scope

Validation is the process of independently assessing the accuracy of a remote sensing product or a model based on RS data by comparing it with reliable reference data. This process focuses on quantifying the errors and reliability of the model's output to iteratively improve the products and underlying algorithms. This enables effective and sound (policy) actions and resource allocation, in our case, with respect to grassland biodiversity monitoring requirements. The validation of remote sensing products relies on several key methods, categorized by the type and availability of the reference data. In this subtask, we explore several validation approaches (direct in-situ validation, indirect validation using ortho-imagery, cross-validation with existing reference layers, and time-series consistency) on several interim datasets from the EU Grassland Watch product suite. Pros and cons are discussed, and some challenges associated with transnational validation are exemplified.

The EU Grassland Watch project is a European initiative dedicated to the effective satellite-based monitoring and management of valuable grassland ecosystems, particularly those within the Natura 2000 network of protected sites (https://ec.europa.eu/eu-grassland-watch/).

Building upon the earlier Copernicus for Natura 2000 (COP4N2K) pilot project, EU Grassland Watch leverages advanced Earth Observation (EO) data from the Copernicus Program (especially Sentinel satellites, combined with Landsat data for historical context) to provide high-resolution time-series information on grasslands. The service aims to empower site managers, national authorities, and the European Commission to proactively assess the status, track changes, and identify pressures on these crucial habitats.

In this Habitat Pilot's subtask, we aim:

- to illustrate the importance of using local reference data and complementary validation strategies in the evaluation of a transnational remote sensing model like EUGW and to identify areas where the model can be improved
- to compare the accuracy of the interim EUGW Typology component in different study areas throughout Europe and indicate some of the model's current limitations
- to assess the potential usability of the EUGW Typology Component for current nature directive monitoring requirements and suggest recommendations for indicator selection and future advancements

Methods and data

EU GRASSLAND WATCH SERVICE

The structure of the EU Grassland Watch product suite consists of four thematic domains, referred to as components. The Land Cover Component provides annual land cover maps based on machine learning classification of the satellite imagery (supported also by other non-EO datasets). It represents the first step in grassland identification and produces a so-called grassland mask defining the area of interest for all subsequent components and data interpretation procedures dealing with grassland characterization. On top of that, the results of the Land Cover Component are also further used for calculating indicators focusing on land cover flows (changes of land cover away from or towards grassland). The grassland characterization part consists of three thematic components: Grassland Typology Component (characterizing grassland from the perspective of biological/habitat type), Grassland Management Component (characterizing grassland from the standpoint of management intensity), and Grassland Productivity Component (characterizing grassland from the perspective of productivity characteristics). Since EU Grassland Watch is conceived as a monitoring tool, one of the main assets is the multitemporal aggregation of the satellite data, resulting in quarterly composites from the pre-Sentinel era (1994-2015) and monthly composites from the Sentinel era (2016-2025)

In the Habitat Pilot project (2024-2025), we evaluated the land cover Component and the Grassland Type Component for several selected study sites and primarily using the most recent available data acquisitions (2023).

The land cover component nomenclature is aligned to a large extent with the Corine Land Cover+ Backbone:

- ➢ LC1 - Urban and built-up areas
- ➢ LC2 - Periodically herbaceous
- ➢ LC3 - Permanent herbaceous
- ➢ LC4 - Broadleaved trees
- ➢ LC5 - Coniferous forest
- ➢ LC6 - Low-growing woody plants

- LC7 - Bare and sparsely vegetated areas
- LC8 - Glaciers and perpetual snow
- LC9 - Water

The Typology Component classifies grassland detections according to EUNIS level 2 classification (Chytrý et al. 2020), namely:

- GT1 - Dry grassland
- GT2 - Mesic grassland
- GT3 - Wet and temporarily wet grassland
- GT4 - Alpine and sub-alpine grassland
- GT5 - Forest clearings and fringes
- GT6 - Inland salt steppes
- GT7 - Sparsely wooded grassland

Method 1: Indirect validation of the grassland mask by means of ortho imagery

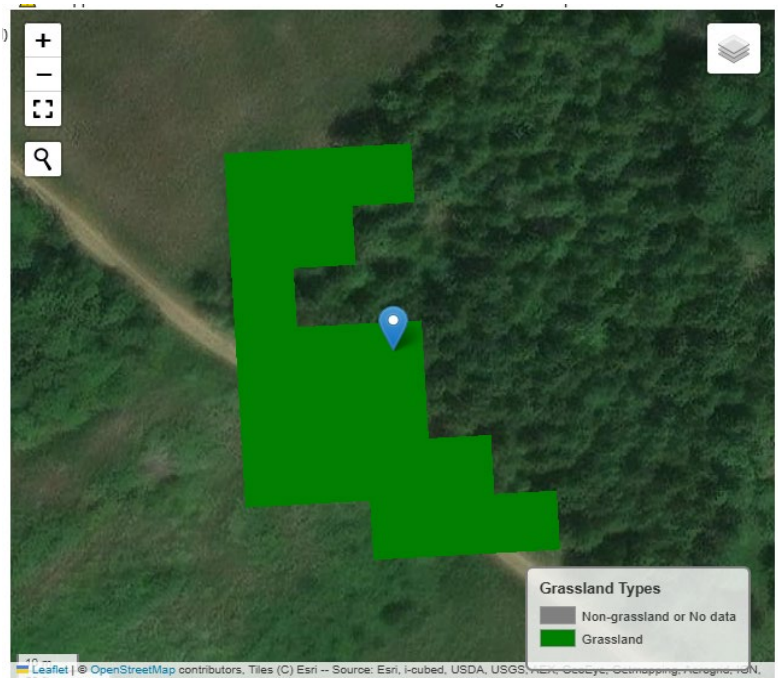
Due to the relative scarcity of accurate and up-to-date ground-truth information on grassland land cover and typology, indirect validation using ortho-imagery was the most obvious approach that could be carried out by multiple partners simultaneously. Although widely applied in the accuracy assessments of RS-products, several important aspects need to be taken into account, which can complicate the interpretation of the results. Some issues we encountered were:

- The scale of the validation is not necessarily the scale at which the end product will be used/produced (e.g., validation on the individual pixel level, while the end product is meant to be used on a larger scale, occurrence of edge effects due to the raster format of the output).
- Definitions used in the model's algorithms are not always compatible with the ones used/intended in the existing land cover typology (e.g., sparsely wooded grassland, mesic grassland, and other habitats insufficiently compatible with the EUNIS typology). Often, there is only partial overlap between habitat types from two typology systems.
- Existing typologies are not always defined unambiguously (e.g., sparsely wooded grassland).
- Geometric distortions can cause a mismatch in the positioning of both the EUGW raster pixels and high-resolution ortho-images, which is especially problematic at edges and borders of different habitats.
- Inter-validator bias originating from pre-existing expertise or the absence of it.
- Inter-validator bias due to different interpretations of the contextual information of a pixel.
- Inter-validator bias due to ad hoc use of additional reference data (LPIS, soil maps, etc).

Specifically, when dealing with land cover identification from ortho-imagery, we need to be aware that:

- ➢ It is often hard to distinguish between intensively managed grass/cropland and more permanent pastures and meadows.
- ➢ It is often hard to distinguish between helophyte-dominated stands (Magnocaricion, Filipendulion, mires and bogs) and (wet) grasslands.
- ➢ It is often hard to distinguish between sparsely wooded grasslands and forest stands or tree plantations.

For our ortho-imagery-based validation cases, a Python notebook was developed and used by the partners to facilitate the ortho-validation of the grassland mask (grassland versus non-grassland) in a harmonized way. In total, 400 random samples were taken from the 2023 EUGW raster files 2023 inside the SAC (Special Areas of Conservation established under the EU Habitats Directive) that were selected a priori. Half, i.e. 200, of the samples were drawn inside class 1 (=grassland) and 200 in class 0 (= non-grassland) from the EUGW raster files. The actual evaluation was carried out in a widget where the selected pixels were being rendered consecutively (see A2_Fig 17). The surrounding 35 pixels were also shown to allow for contextual interpretation when necessary. Temporally aligned ortho-imagery for 2023 was loaded into the scripts by each individual partner. Points were assigned either a TP (true positive), FP (false positive), TN (true negative) or FN (false negative) label (see A2_Fig 18).



A2_Fig 17 Screenshot of the evaluation widget used for labelling the randomly sampled pixels. The pixel indicated with a marker is the one under evaluation. Model output is provided for the surrounding 35 pixels (of which 18 are classified as "grassland") to allow for contextual interpretation.

Prior to the actual evaluation, several decision rules were discussed and agreed upon by the validators:

- ➢ The standard level of assessment of a selected pixel is the centroid.
- ➢ In the case of a mixed land cover pixel, we look at the majority of the pixel.

- ➢ In some cases, the neighbouring pixels may assist in evaluation (e.g. grassland borders).
- ➢ Assign one of the following labels.
- ➢ When unsure, assign the label 'uncertain' and document the case.

For EUNIS level 2 grassland classifications, we referred to the definitions of the EUNIS habitat version 2021 of group R (https://eunis.eea.europa.eu/habitats-code-browser-revised.jsp) and the crosswalk with EuroVegChecklist (Schaminée et al. 2018) for local syntaxonomical interpretations.

Some additional points of attention were discussed:

- EUNIS Level 2 Mesic grassland: this EUNIS habitat type covers a wide range of well-drained pastures and hay meadows from intensively used, species-poor, fertilized grasslands to species-rich, traditionally managed grassland with very rare plant species

-EUNIS Level 2 Wet and seasonally wet grassland: on ortho-imagery, this EUNIS class is hard to differentiate from fen, mire, or bog vegetations. It could be useful to label as 'uncertain' or 'false' when a wetland, instead of a grassland, is suspected

-Alpine and sub-alpine grasslands: in ecological terms, a wide habitat category including all open vegetation in alpine or sub-alpine zones that are not included in EUNIS classes Q (wetlands), S (heathlands), or U (Sparsely vegetated)

-Dry grasslands: these include various types of grassland habitats, such as sparsely vegetated pioneer grassland on sandy or rocky soils, dunes, among others.

From the 400 labelled samples, the following standard metrics were calculated:

Overall Accuracy A= (TP + TN) \ (TP + TN + FP + FN)

providing a measure of the proportion of total correct predictions.

Precision P = TP\ (TP + FP)

a measure of the model's performance for predicting 'positive' values: of all the instances the model predicted as positive, how many were actually positive?

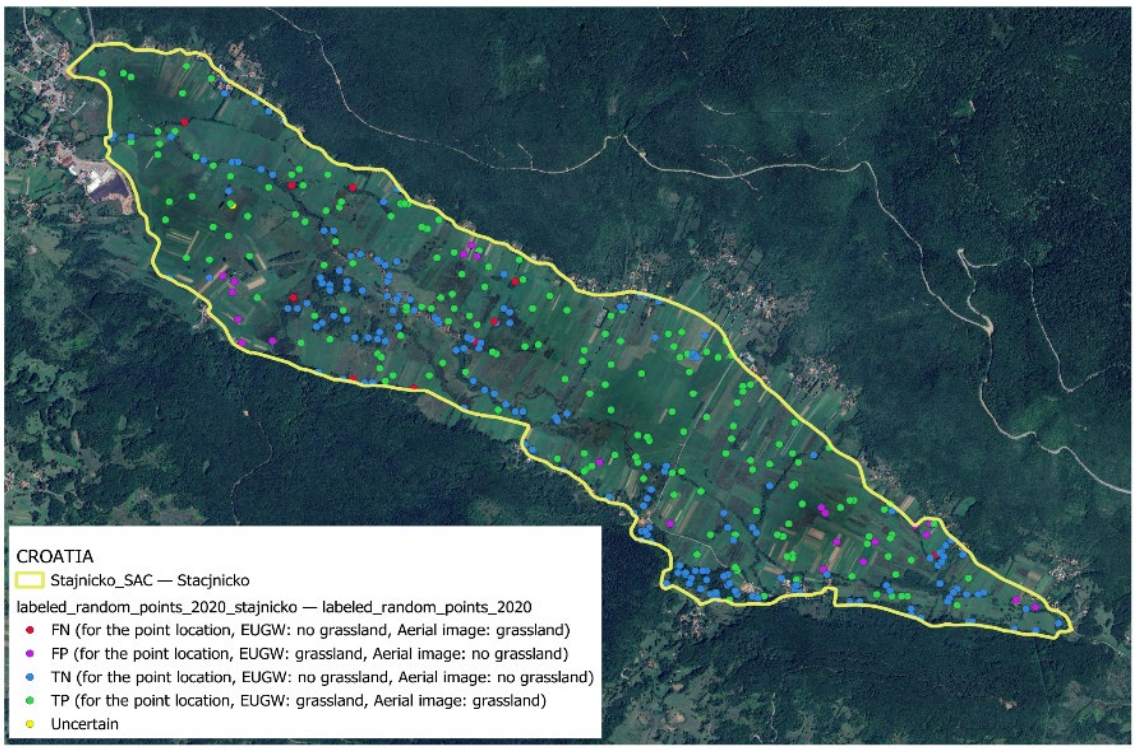
Recall R= TP\ (TP + FN)

a measure of the model's ability to find all positive samples: of all the instances that were actually positive, how many did the model correctly identify?

F1 Score F1 = 2.Precision.Recall/ (Precision + Recall) = 2xTP/ (2xTP + FN + FP)

the harmonic mean of Precision and Recall, especially useful when there is an uneven class distribution (class imbalance).

All of the above values were corrected for the actual proportions of grassland and non-grassland pixels present in the test population.



A2_Fig 18 Example from SAC Stajnicko, Croatia

Results

Table 3 summarizes the (corrected) accuracy metrics of the EUGW grassland mask in different SAC areas throughout the EU. The overall accuracy of the model was high, to very high, in correctly classifying grassland versus non-grassland. Specific differences are discussed at the test site level below. More general misclassifications occur in part due to edge effects inherently related to the raster output of the model. As a result, there are often inaccuracies in delineating the grassland polygon, but this kind of error is believed to be of lesser concern for most applications. Another general finding was that the sparsely wooded class of the grassland mask seems inconsistent between different areas and is in general, overfitted by the model. Tree lines, hedgerows, and scattered trees are correctly treated by the model as grassland, but often classified as sparsely wooded grassland (GT7), which, depending on the definition of the class being applied, is not always the case. For example, a significant area of grassland with partial tree cover in Swedish and Czech SACs is classified as grassland, while in other areas this is not the case. The amount of tree cover, with regard to habitat definition but also grassland quality status (tree and shrub encroachment), is an essential aspect and requires dedicated attention in the future improvements of the EUGW model. Another important observation that was made in several areas is the classification of high-productivity grass croplands as (mostly mesic) grassland. Although not necessarily in contradiction with EUNIS class definitions, this kind of ‘grass cropland’ is of minor interest from a nature conservation standpoint.

Finally, from A2_Table 3, it becomes apparent that applying an EU-wide grassland detection model, even when harmonized on the biogeographical level, inevitably works much better in certain areas

than in others. It is, however, promising to observe lower accuracies mainly occurring in areas with very specific and unique grassland characteristics (see also the test sites discussion below).

A2_Table 3 Summary of the accuracy metrics for the evaluated SACs (values indicated with an asterisk are uncorrected for grassland/non-grassland proportions)

Country	SAC	#uncertain labels	Overall accuracy	Precision	Recall	F1 score
Sweden	Stora Alvaret	1	0.76*	0.84*	0.58*	0.69*
Sweden	Tromtö-Almö	36	0.87*	0.90*	0.86*	0.88*
Sweden	Södra-Stensjö	65	0.73*	0.85*	0.73*	0.79*
Czechia	Milovice-Mlada	60	0.97	0.91	1.000	0.95
Flanders	Semois	18	0.94	0.92	0.97	0.94
Flanders	Voeren	7	0.92	0.87	0.97	0.92
Denmark	Mols Bjerge	12	0.75	0.92	0.63	0.75
Croatia	Livade uz Bednju II	0	0.96	0.95	0.97	0.96
Croatia	Stajnicko	1	0.90	0.98	0.99	0.94
Slovakia	Orava	0	0.93	0.87	0.98	0.92
Slovakia	Latorica	2	0.92	0.82	0.94	0.87
Bolzano	Valle di Funes	10	0.93	0.97	0.85	0.90
Bolzano	Prati dell'Armentara	5	0.96	0.98	0.94	0.96

Site level discussion

1. SAC CZ0214006 Milovice-Mladá (Czechia)

EUGW blocks evaluated: CZ_CON_CON_75111_20230101_20231231_GTYH_CLASS

A relatively high number of pixels were labelled as 'uncertain' (60) from the ortho-imagery, mainly because of the ambiguous habitat definition regarding tree encroached grassland. False positives were identified as small non-grassland elements within larger grassland polygons, e.g. military tracks, sparsely vegetated parts or heathland vegetation. No false negatives were encountered.

2. SAC SKUEV0006 Latorica (Slovakia)

EUGW blocks evaluated: HR_ALP_MDM_76353_20230101_20231231_LCMH_CLASS, HR_ALP_MDM_76353_20230101_20231231_LCMH_CLASS

Some pixels (13) were labelled as uncertain due to possible confusion with arable fields. False positives were mainly edge effects, but confusion with arable fields and wetland (tall herb) vegetation was observed. Again, no false negatives were observed, but the area has a very high ratio of grassland to non-grassland.

3. SAC SKUEV0243 Horná Orava (Slovakia)

EUGW blocks evaluated: SK_ALP_CON_81737_20230101_20231231_GTYH_CLASS, SK_ALP_CON_81736_20230101_20231231_GTYH_CLASS, SK_ALP_CON_82038_20230101_20231231_GTYH_CLASS

Some confusion of mainly dry grasslands with arable lands (barley, weed), forest clearings, or grass cropland (clover/alfalfa) was observed. Fallowing is part of the local agricultural practice in this area and may account partly for this kind of misclassification. When looking qualitatively into the typology layer, only dry, mesic (and some wet and temporarily wet) grassland were classified by the model. This accurately reflects the current situation where a subalpine grassland site has been abandoned by agriculture and is overgrown with dwarf pine and scrubs (*Vaccinium* spp.). Also, some overfitting of sparsely wooded grassland (f.e. gardens, tree lines, and forest edges) and forest clearings (currently still trees present) was noted.

4. SAC IT3110026 Valle di Funes (Italy)

EUGW blocks evaluated: IT_ALP_ALS_70624_20230101_20231231_LCMH_CLASS, IT_ALP_ALS_70924_20230101_20231231_LCMH_CLASS, IT_ALP_ALS_70925_20230101_20231231_LCMH_CLASS, IT_ALP_ALS_70623_20230101_20231231_LCMH_CLASS

False negatives (8) due to edge effects and tree cover, but also in areas with mixed bare soil and grassland patches. Only one sample pixel remained undetected by the model due to shadows. False positives were tree-covered pixels near forested areas. Uncertain labels were assigned to tree-encroached areas and some rocky surfaces within a grassland matrix. Most likely, these pixels could, depending on the definition, be labelled as true positives since they are part of a larger grassland matrix.

5. SAC IT3110048 Prati dell'Armentara (Italy)

EUGW blocks evaluated: IT_ALP_ALS_70925_20230101_20231231_LCMH_CLASS, IT_ALP_ALS_70924_20230101_20231231_LCMH_CLASS

Misclassifications are in the same order of magnitude as in the neighbouring Valle di Funes, with false negatives (5) due to edge effects and one pixel with shadow. Some pixels were labelled uncertain due to edge effects of tree groups inside a grassland. Also, edge effects and rocky soil caused a couple of false positives.

6. SAC SE0410216 Södra stensjö (Sweden)

EUGW blocks evaluated: SE_BOR_CON_74765_20230101_20231231_GTYH_CLASS

Although not specifically mentioned in the Standard Data Forms, Habitat directive Annex 1 habitat type 9070, Fennoscandinavian wooded pastures, is present in this area and accounts for the majority of misclassifications by the model. As in other SACs, edge effects are also adding to the false negatives and positives. Interpretation of the ortho-imagery in this case was combined with field validation, drone imagery, and expert knowledge on the actual vegetation type. Thus, it should be taken into account that from ortho-imagery alone it is often not possible to discriminate between sparsely wooded grassland types and forest stands or plantations. As a result, an important share

of the sampled pixels was labelled uncertain (60). It was also observed that isolated buildings and cropland were misclassified as grassland.

7. SAC SE0330176 Stora Alvaret (Sweden)

EUGW blocks evaluated: SE_CON_CON_76269_20230101_20231231_GTYH_CLASS, SE_CON_CON_76268_20230101_20231231_GTYH_CLASS, SE_CON_CON_75969_20230101_20231231_GTYH_CLASS, SE_CON_CON_75969_20230101_20231231_GTYH_CLASS, SE_CON_CON_75967_20230101_20231231_GTYH_CLASS

Due to the complexity of the Stora Alvaret area, a number of specific misclassifications seem to occur in the model. Alvar is a peculiar limestone landscape consisting of grassland, heathland, wetland, scrubland, and rocky habitats. A clear definition of what is considered grassland and non-grassland is therefore crucial. For the evaluation, limestone pavements were considered non-grassland, whereas patches of shrubs and heathland were considered part of the grassland definition. In practice, it remains very challenging to discriminate between Annex 1 habitat 8240 Limestone pavements and Annex 1 grassland types 6110 Rupicolous calcareous or basophilic grasslands of the Alysso-Sedion albi, 6210 Semi-natural dry grasslands and scrubland facies on calcareous substrates (Festuco-Brometalia) (* important orchid sites), as well as Annex 1 type 6280 Nordic alvar and Precambrian calcareous flatrocks even in the field, because these types often occur in an interspersed mosaic.

Calculated accuracy for grassland detection in Stora Alvaret is thus not surprisingly amongst the lowest for the evaluated areas. There is a significant proportion of false negatives (72) due to unharmonized definitions (low woody is sometimes part of the Annex 1 grassland definition), overfitting of bare soil, and increased effect of edge and scale due to the fine-scale interspersion of Alvar subtypes. False positives were far less identified, but can be attributed to the same causes. It was also observed in this SAC that there is an important misclassification of bare and sparsely vegetated soil with sealed/urban land cover types.

8. SAC SE0410042 Trömtö Almö (Sweden)

EUGW Blocks evaluated: SE_BOR_CON_74766_20230101_20231231_GTYH_CLASS

Interpretation of the grassland detection in Trömtö Almö is hampered again by the relatively high presence of Fennoscandinavian wooded pastures and meadows (Annex 1 habitat codes 9070 and 6530). Without ground truth data on the actual vegetation occurring, it is almost impossible to discriminate between sparsely wooded grassland or non-grassland areas with high amounts of tree cover. It is therefore very unlikely that the model can deliver high accuracy on these specific classes of grasslands.

9. SAC HR2000634 Stajničko polje (Croatia)

EUGW Blocks evaluated: HR_ALP_MDM_76354_20230101_20231231_LCMH_CLASS, HR_ALP_MDM_76353_20230101_20231231_LCMH_CLASS

Stajničko polje is a grassland-dominated SAC with an important area of Molinia meadows on calcareous, peaty or clayey-silt-laden soils (Molinion caeruleae) - Annex 1 code 6410. Grassland detection was very reliable, with only a limited number of false negatives entirely attributable to edge

effects. False positives were a little higher due to confusion with arable land. In some rare cases, misclassifications seem to be related to high-intensity grass croplands. When needed, the LPIS layer was used to help distinguish grasslands from arable land.

10. SAC HR2001409 Livade uz Bednju (Croatia)

EUGW blocks evaluated: HR_CON_MDM_77548_20230101_20231231_LCMH_CLASS

Livade uz Bednju II is a Natura 2000 area of wet, species-rich grassland where hay-making has been historically practiced. Typical habitats in the area are tall-herb communities of wet meadows, Medio-European submontane hay meadows and water-fringing reedbeds and tall helophytes. In the last habitat map, a big proportion of the area was classified as agricultural land. The EUWG model classified some unmanaged parcels of land, showing clear signs of succession, as grasslands. This interpretation was accepted during the validation, as these areas can still be considered grasslands in a broader ecological sense. However, it should be noted that their floristic composition is significantly degraded compared to the typical habitat structure and partial succession towards wet tall-herb communities is probably underestimated.

11. SAC BE34057C0 Marais de la Haute Semois et Bois de Heinsch (Belgium)

EUGW Blocks evaluated: BE_CON_CON_61570_20230101_20231231_GTYH_CLASS, BE_CON_CON_61871_20230101_20231231_GTYH_CLASS, BE_CON_CON_61872_20230101_20231231_GTYH_CLASS

Several sampled pixels were labelled as uncertain because the suspected land cover as non-grassland wetland, or very intensively managed grass cropland.

False negatives were entirely attributable to edge effects, and false positives also resulted from model confusion with non-grass wetland and abandoned, tree-encroached grassland.

12. SAC BE2200039 Voerstreek (Belgium)

EUGW Blocks evaluated: BE_CON_CON_62165_20230101_20231231_GTYH_CLASS

Some misclassification of cropland (maize) and wetland tall herb communities was observed and accounted for, along with edge effects, for a limited number of false positives. Some small grassland polygons, often linear, remained undetected.

13. SAC DK (Denmark)

EUGW blocks evaluated: DK_CON_CON_68746_20230101_20231231_GTYH_CLASS

A rather large amount of false negatives were reported for this area mainly involving less dense stands of broom (Cytisus spp.) that occur scattered throughout larger blocks of grassland. Some false positives were also observed where the model detected (dry and sparsely wooded) grassland on pixels with urban constructions.

Method 2: Cross-validation interim EU Grassland Watch typology and land cover with local reference datasets

SAC BE34057C0 Marais de la Haute Semois et Bois de Heinsch (Belgium)

The 2021 wall-to-wall Waleunis polygon map for SAC BE34057C0 Marais de la Haute Semois et Bois de Heinsch, kindly provided by the Department for the study of the natural and agricultural Environment of the Walloon Government in Belgium, was used to assess the accuracy of the typology output of the EUGW model at the EUNIS level 2 classification level. Waleunis' typology is the local adaptation of the European EUNIS typology for the Walloon region, describing all terrestrial and aquatic biotopes occurring in the territory. As the first step, the map units were reclassified to fit EUNIS level 1 (Waleunis L1) and subsequently grassland and wetland groups to EUNIS level 2 (Waleunis L2). In the next step, polygon maps were rasterized and aligned with the intersecting EUGW typology blocks in order to perform a pixel-by-pixel comparison. Analysis was carried out both in R (using terra and caret packages) as well as in QGIS (using the SCP plugin) for the visualization of the discrepancies.

Results

From the level 1 comparison, the relative numbers show (see A2_Fig 19) :

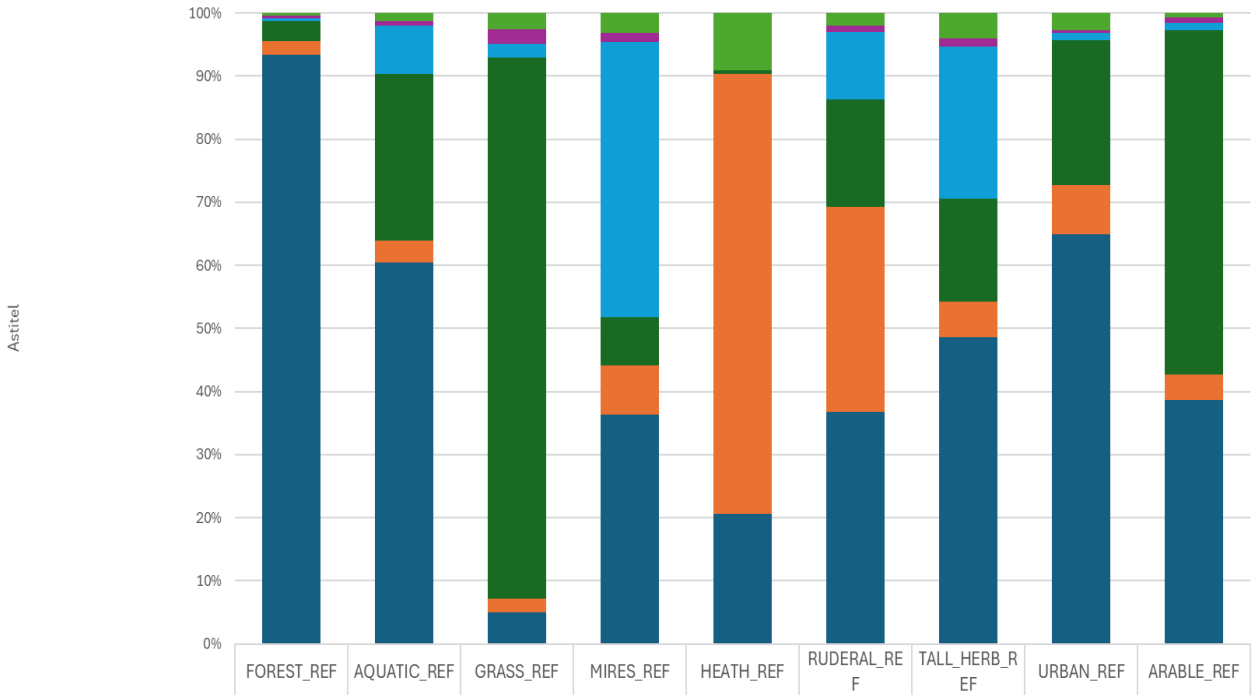
-some misclassification of EUGW class mesic grassland with arable land (grasscrops), urban areas (gardens), and to a lesser extent also tall herb communities (non-grassland wetland) and aquatic vegetations.

-dry grassland EUGW class is primarily detected in areas mapped as heathlands or ruderal vegetation.

-wet and temporarily wet EUGW class is to a large extent detected in mires, aquatic vegetations, and tall herb communities

-sparsely wooded grasslands and forest clearings are detected in a broad range of reference biotopes

When accuracy is expressed in absolute surface numbers, only 4.7% of the reference grassland area remained undetected by the model. Mesic grassland was classified for 91% within the grassland reference stratum, dry grassland for 36%, wet grassland for 28% and sparsely wooded for 70%. The forest clearing class was not calculated because this could not be matched with reference biotope typology from the Waleunis maps.



	FOREST_REF	AQUATIC_REF	GRASS_REF	MIRES_REF	HEATH_REF	RUDERAL_REF	TALL_HERB_REF	URBAN_REF	ARABLE_REF
SPARSELY WOODED_EUGW	4,1	0,2	26,8	2,6	0,1	0,1	2,6	1,2	0,3
FOREST CLEARINGS_EUGW	4,0	0,1	23,1	1,2	0,0	0,1	0,8	0,2	0,5
WET GRASS_EUGW	3,6	1,1	22,4	36,1	0,0	0,7	15,3	0,6	0,6
MESIC GRASS_EUGW	30,7	3,8	871,1	6,3	0,0	1,1	10,4	10,3	28,0
DRY GRASS_EUGW	20,5	0,5	21,9	6,4	1,1	2,1	3,6	3,5	2,1
NA_EUGW	886,6	8,7	50,5	30,0	0,3	2,3	30,9	29,2	19,9

A2_Fig 19 Results from pixel-by-pixel comparison of EUGW classes with L1 Waleunis vegetation map for SAC BE34057C0 Marais de la Haute Semois et Bois de Heinsch expressed as relative proportions (stacked bars) and absolute number of hectares (table).

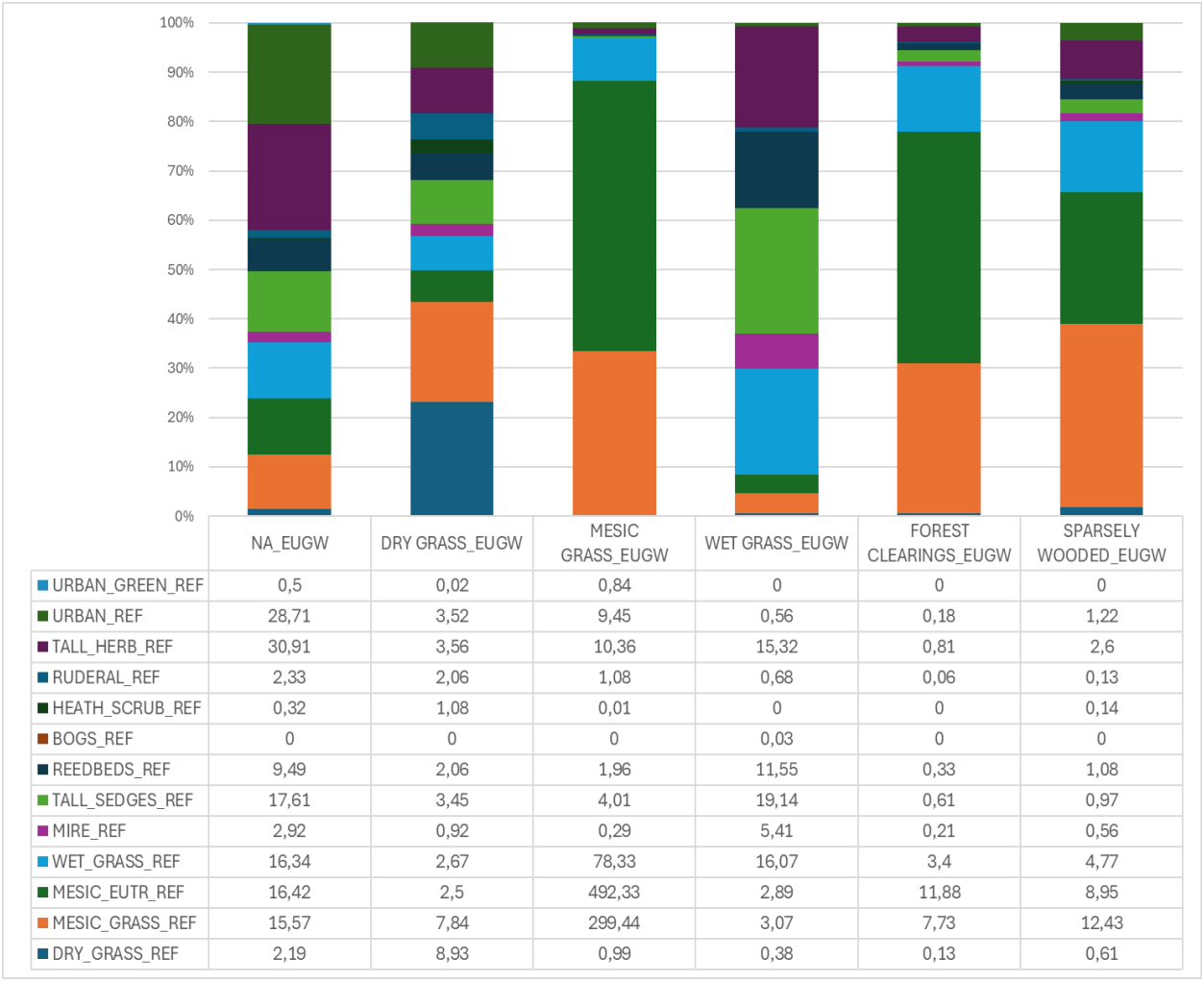
Results from level 2 comparison provide further insight into model confusion with regard to the lower-level reference layer classes (see A2_Fig 21). Within the reference class of mesic grasslands, Waleunis allows to differentiate between species-poor (MESIC_EUTR) and species-rich (MESIC_GRASS) mesic types. In more detail, the results show that:

-the EUGW model detects 68% of dry reference grassland as EUGW class dry grassland, but there is significant confusion with mesic grassland, heathland, and ruderal reference biotopes.

-the EUGW model detects 13% of the wet reference grassland as EUGW class wet and temporarily wet grassland, and shows significant confusion with non-grassland reference wetlands (mires, tall sedges, and reedbeds)

-the EUGW model correctly classifies 86,5% and 92% of the mesic reference class, respectively, the species-rich and species-poor subclass.

-the EUGW class sparsely wooded grassland is to a large extent detected within wet and mesic grassland, but a significant portion is also identified as forest in the reference data.



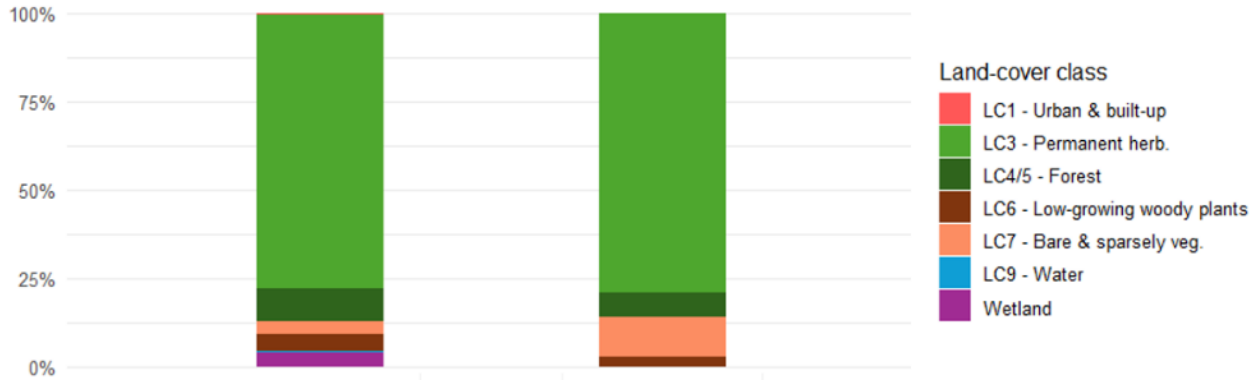
	NA_EUGW	DRY GRASS_EUGW	MESIC GRASS_EUGW	WET GRASS_EUGW	FOREST CLEARINGS_EUGW	SPARSELY WOODED_EUGW
URBAN_GREEN_REF	0,5	0,02	0,84	0	0	0
URBAN_REF	28,71	3,52	9,45	0,56	0,18	1,22
TALL_HERB_REF	30,91	3,56	10,36	15,32	0,81	2,6
RUDERAL_REF	2,33	2,06	1,08	0,68	0,06	0,13
HEATH_SCRUB_REF	0,32	1,08	0,01	0	0	0,14
BOGS_REF	0	0	0	0,03	0	0
REEDBEDS_REF	9,49	2,06	1,96	11,55	0,33	1,08
TALL_SEDGES_REF	17,61	3,45	4,01	19,14	0,61	0,97
MIRE_REF	2,92	0,92	0,29	5,41	0,21	0,56
WET_GRASS_REF	16,34	2,67	78,33	16,07	3,4	4,77
MESIC_EUTR_REF	16,42	2,5	492,33	2,89	11,88	8,95
MESIC_GRASS_REF	15,57	7,84	299,44	3,07	7,73	12,43
DRY_GRASS_REF	2,19	8,93	0,99	0,38	0,13	0,61

A2_Fig 20 Results from pixel-by-pixel comparison of EUGW classes with L2 Waleunis vegetation map for SAC BE34057C0 Marais de la Haute Semois et Bois de Heinsch expressed as relative proportions (stacked bars) and absolute number of hectares (table).

SAC IT3110029 Parco Naturale dello Sciliar-Catinaccio

Similar to Flanders, a local wall-to-wall polygon map for habitats in the Parco Naturale dello Sciliar-Catinaccio from 2013 was used to assess the accuracy of the output from the EUGW model. At the time of the Habitat Pilot, the EUGW grassland typology layer was not yet available for the area. Thus, the local habitat map was compared with the output of the EUNIS Level 1 Land Cover component data.

Like in Flanders, the classifications of the local polygon map (Wilhalm et al., 2022) were first aligned with the Land Cover component data (corresponding to the classifications used in CLC+ Backbone, see 'Methods & Data') and the polygon map rasterized and aligned with the EUGW data in R using the packages terra, exactextractr and caret. Finally, the inconsistently classified pixels were reviewed using a local orthophoto from 2023 in QGIS. The results from the pixel-by-pixel analysis are visualized in A2_Fig 21.



	GRASS_EUGW (LC3)	GRASS_REF (LC3)	
LC1_REF	10	0	LC1_EUGW
LC3_REF	1461	1461	LC3_EUGW
LC4/5_REF	169	122	LC4/5_EUGW
LC6_REF	87	50	LC6_EUGW
LC7_REF	74	211	LC7_EUGW
LC9_REF	7	0	LC9_EUGW
Wetland_REF	77		

A2_Fig 21 Results from pixel-by-pixel comparison of the area classified as grassland (LC3 - Permanent herbaceous) in the EUGW grassland mask (left) and a local habitat map (right) for Parco naturale dello Sciliar-Catinaccio in Bolzano expressed as relative proportions (stacked bars) and absolute number of hectares (table).

The results show some confusion between forest and grassland habitats in the two maps:

- Areas classified as grassland in the local habitat map and forest in the EUGW land cover component are primarily sparsely wooded grasslands.

- Areas classified as forest in the local habitat map and grassland in the EUGW land cover component are primarily forested areas that have been clear-cut.

Additionally, the comparison showed that some high-alpine, low-productive grasslands, primarily in areas mixed with bare rock surfaces, are classified as ‘Bare & sparsely vegetated’ in the EUGW land cover component.

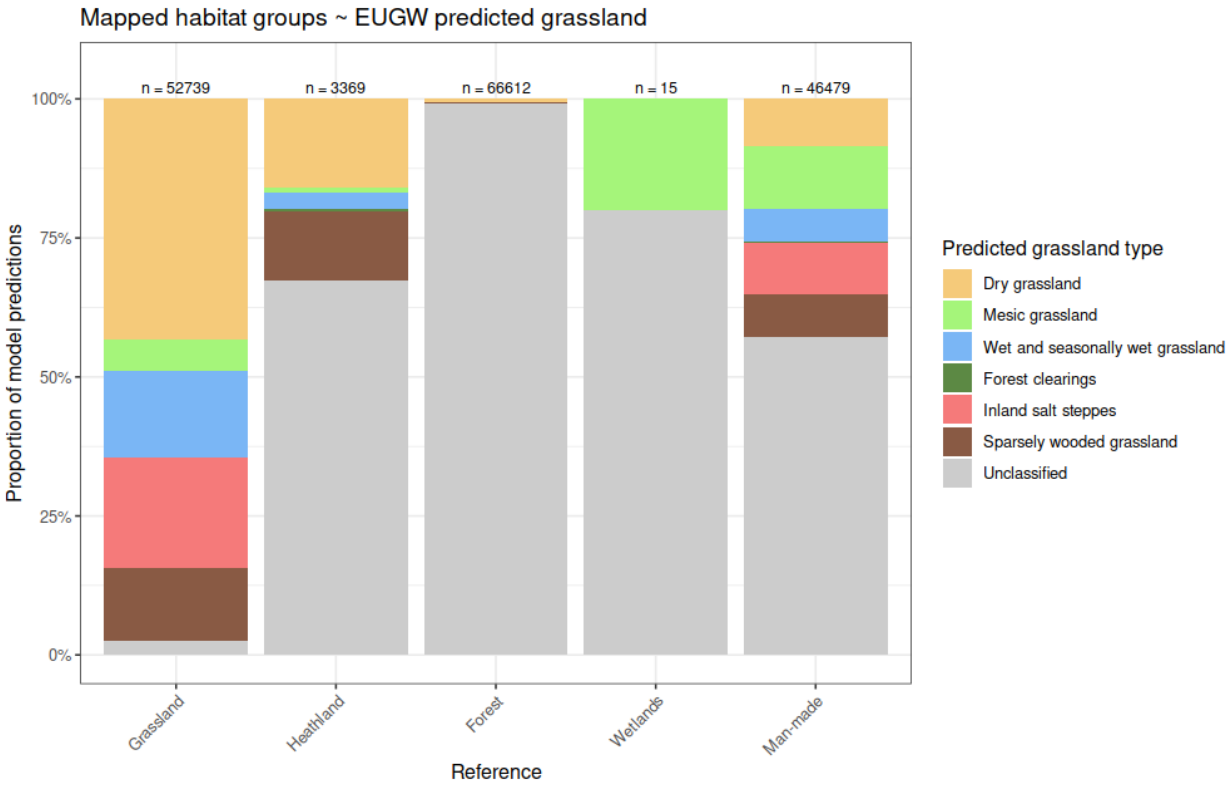
Finally, all areas classified as ‘Wetland’ in the local habitat map are shown as ‘Grassland’ in the EUGW land cover map. Given that some of these grass-dominated wetlands are classified as degraded wetlands and might more accurately be considered wet grasslands—especially since the habitat map is over 10 years old—the extent of misclassification between wet grassland and non-grass wetland remains uncertain.

SAC CZ0214006 Milovice-Mladá (Czechia)

For the Czech site Milovice–Mladá, the analysis followed a similar approach as for the two previously described locations. The local wall-to-wall habitat map was used to assess the EUGW model output, but first had to be rasterized to match the spatial resolution and extent of EUGW in order to enable a pixel-by-pixel comparison. The Czech habitat classification system was subsequently translated into the EUNIS 2021 scheme.

The results show that the model correctly identifies the vast majority (97.4%) of the area field mapped as “Grassland”, which is consistent with the indirect validation based on orthophotography (see

Method 1). Most of the reference “Grassland” pixels were classified by the EUGW model as Dry grassland (43.4%), although a considerable proportion was also assigned to Inland salt steppes (19.8%), Wet and temporarily wet grasslands (15.6%), and Sparsely wooded grasslands (13.1%). However, the model also classified a substantial share of Man-made habitats (42.8% in total) as “Grasslands”, which is likely due to confusion with agricultural land, gardens, and golf courses. Another notable mismatch is the misclassification of 32.6% of field-mapped “Heathlands”, probably caused by an edge effect associated with linear shrub formations.



A2_Fig 22 Results of the pixel-by-pixel comparison of the local habitat mapping layer with the EUGW grassland typology output for the Czech site Milovice–Mladá (CZ0214006). Stacked bars represent the total area mapped as a given EUNIS category, while the colours represent the grassland type classified by the EUGW model. The numbers above the bars indicate how many pixels were mapped in the field as the given EUNIS category.

Method 3: Direct in-situ validation of the interim EUGW typology component

The Typology Component (EUNIS level 2 classification) was evaluated in a number of study areas from the Continental zone for which datasets were available during the project. Table X2 shows an overview of the field surveys carried out. Guidelines were provided on sample size and selection, but were not always applied in the same manner by the partners due to time constraints. EUNIS L2 class definitions and interpretations were discussed a priori by all participating partners. Field work was carried out during the (late) summer of 2025.

A2_Table 4 Summary information of the field surveys conducted in the Habitat pilot EUGW subpilot.

Country/partner	SAC	Environmental Zone	Inspected class	Period of sampling
MoE-Denmark	Mols Bjerge	Continental	Dry grassland	Sep 2025
MoE-Denmark	Mols Bjerge	Continental	Mesic Grasslands	Sep 2025
SEPA-Sweden	Södra-Stensjö	Continental	Sparsely wooded grassland,	July 2025
EURAC-Bolzano	Parco naturale dello Sciliar-Catinaccio	Alpine	Land cover component	July 2025
AOPK-Czech republic	Milovice-Mlada	Continental	Dry grassland	August 2025

SAC DK00DX300 Molslaboratoriet, Mols Bjerge (Denmark)

To validate the EU Grassland Watch predictions of the grassland type, 100 points were selected randomly in the model-predicted dry grassland stratum and 100 points in the model-predicted mesic grassland stratum. Subsequently, using ESRI Field Maps on a smartphone, each point in the field was surveyed, and it was determined whether or not the habitat was dry respectively mesic grassland or not. The model predictions were validated for dry or mesic grasslands by standing in the center of the sampled 10 x 10 m quadrat. Based on the composition of the plant communities in the quadrat, it was determined whether the habitat type of the 10-m plot was a grassland and whether it was dry or mesic. Quadrats with more than 50% coverage of shrubs and trees were not considered grasslands.

In our assessments of the vegetation, we used EUNIS level 2 grassland definitions and used the vascular plant composition as indicative of the dryness/mesicness. Dry quadrats were dominated by low-growing perennial herbaceous species, including several species with traits adapted for dry conditions, such as *Hieracium pilosella* and *Hypochaeris radicata*. Mesic grasslands were also dominated by low-growing perennial herbaceous species, but in these, there were typically several species typical for moderately wet conditions, such as *Juncus effusus*, *Cirsium palustre*, and *Potentilla erecta*.

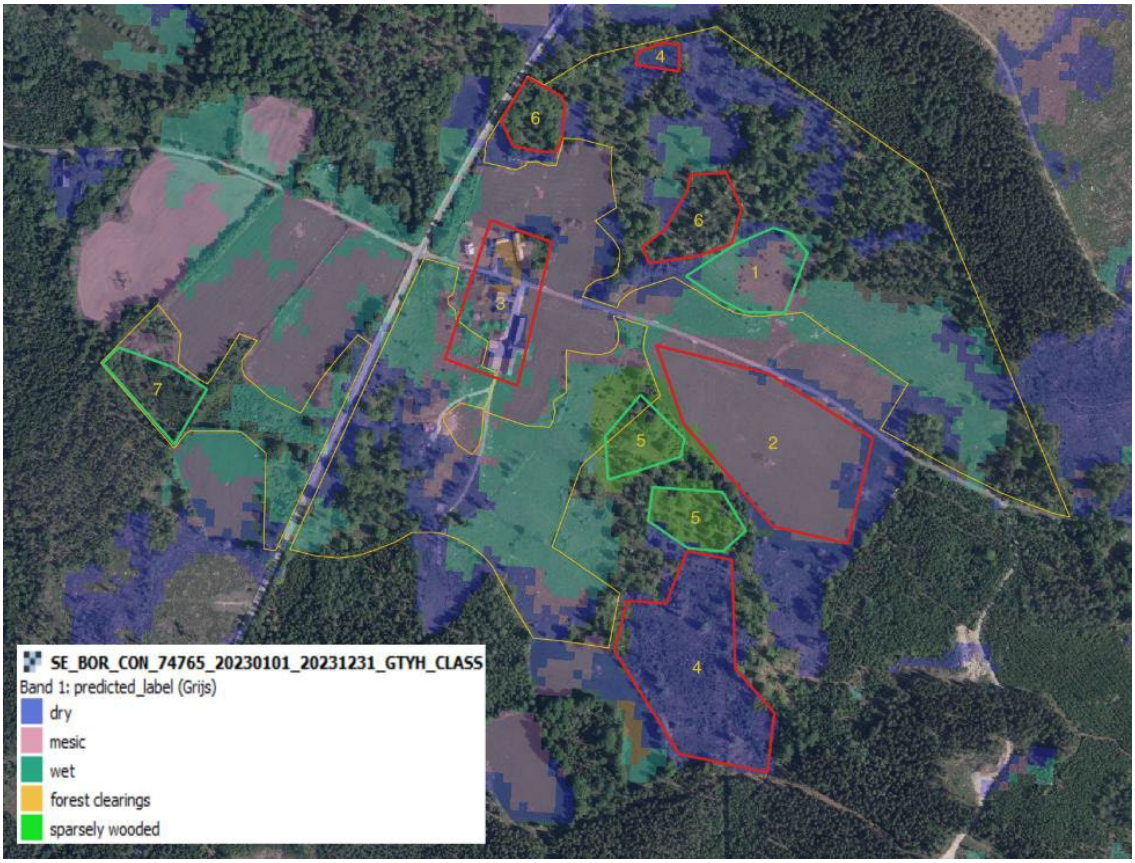
The classification accuracy for dry grassland was measured as 41% and for mesic grassland 10%. The high proportion of misclassifications was primarily attributed to tree and shrub encroachment (*Prunus* sp., *Cytisus* sp., *Rubus* sp.) for both classes. This finding seems contrary to the results from the ortho-interpretation exercise in method 1 for this area, where shrub-covered grassland was rather underdetected. This illustrates the importance of unambiguously defining the scope of a certain habitat class to allow for a robust validation process. This case study also illustrates the need for

strict definitions. For example, to what extent woody plant cover in grasslands is considered part of the grassland ecotope or not.

SAC SE0410216 Södra Stensjö (Sweden)

A qualitative in-situ evaluation of EUGW Typology output was performed for SAC Södra Stensjö in Sweden. A number of observations were made and are illustrated in A2_Fig 23:

1. The model correctly classifies certain open Annex I grasslands as grasslands
2. It also classifies regularly plowed cropland as grassland
3. Buildings sometimes cause confusion. This area of buildings is classified as forest fringe and dry grassland
4. Clear-cut forests are classified as grassland classes even though the field vegetation has not changed. This was evident in both areas with or without access to grazing animals.
5. Sometimes the model correctly identifies sparsely forested areas
6. These areas were typically grazed oak forests, which were not recognized as sparsely forested areas, even if they most likely would fall into the Annex I class Fennoscandinavian wooded pastures.
7. Previously grazed grassland that is now covered by dense shrubs was correctly classified as not grassland



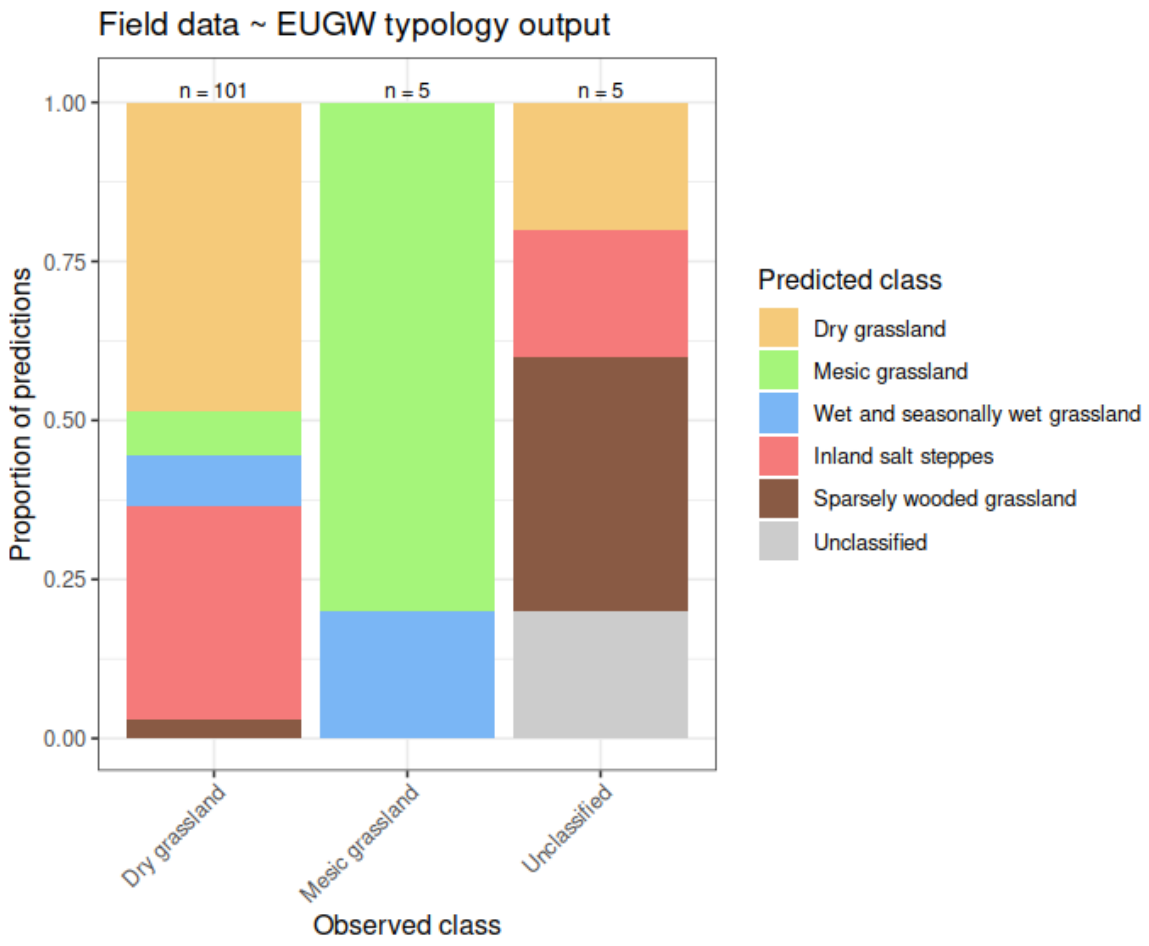
A2_Fig 23 In-situ inspected areas in Södra Stensjö, Sweden.

SAC CZ0214006 Milovice-Mladá (Czechia)

Fieldwork at the Milovice – Mladá site focused on validating the Dry grasslands category. Since the field survey was conducted before the delivery of the EUGW typology product, the Habitat mapping layer of Czechia (Vrstva mapování biotopů ČR) was used to stratify the validation points. This layer is based on a national classification system, which was translated into the EUNIS 2021 scheme used in the EUGW product, following the Habitats catalogue of the Czechia (Chytry et al., 2010) and additional reference materials. At each survey location, a 10 by 10 metre plot – corresponding to a single Sentinel-2 pixel – was assessed. Based on the recorded plant species, the plot was assigned to the appropriate EUNIS Level 2 category. We also collected additional site characteristics regarding the observed grassland management.

Out of the 111 surveyed locations, 101 were confirmed in the field as Dry grassland (90.9%). Additionally, five locations were classified as Mesic grassland (4.5%), three as no-grassland (2.7%), and two as sparsely wooded grassland (1.8%). The sparsely wooded grassland category was subsequently treated as non-grassland, as there is no vegetation type corresponding to this EUNIS category in Czechia, according to the classification translation described above.

The comparison of field data with the EUGW typology output shows that the model correctly identified the Dry grassland category in almost half of the cases (48.5%). However, in one third of the cases (33.7%), this category was incorrectly classified as Inland salt steppes, a habitat type that does not occur in the study area according to multiple previous expert surveys. It should be noted that this error was spatially concentrated mainly in the extensively grazed enclosures with large herbivores (*Bison bonasus*, "*Bos primigenius*", *Equus ferus*) and in the mown areas of the former military airfield and might be related to the presence of small scattered patches of bare soil.



A2_Fig 24 Proportional distribution of EUGW typology predictions for three groups of field-observed habitat categories at Milovice - Mladá site (CZ0214006).

SAC IT3110029 Parco naturale dello Sciliar-Catinaccio (Italy)

For the field validation in Bolzano, Parco naturale dello Sciliar-Catinaccio was selected for its high proportion of grassland habitats and the availability of a detailed local reference habitat map. Sciliar-Catinaccio contains large alpine Annex 1 habitat types, particularly meadows including, e.g., sites with Alpine calcareous grasslands (6170), Mountain Hay Meadows (6520), and Species-rich Nardus grasslands (6230). Two hundred random points were generated in the more accessible parts of the protected area, in areas classified as Alpine Grasslands from the local reference habitat map.

In July 2025, 157 out of the 200 sites were field validated, whereas the remaining 43 random points could not be visited due mostly to the steepness of the terrain. At each visited site, the dominating habitat type for the 10x10 meter area (corresponding to one Sentinel-2 pixel) surrounding each point was classified, based on the present plant communities at the site. Out of the 157 validated points, ten points (6%) were assigned different EUNIS level-2 classifications in the field than in the local habitat map. Out of the ten points, five were, e.g., classified as “bare & sparsely vegetated” (all however in the border zone with grassland areas), and three points were classified as Mesic Grassland. Since EUGW Typology is currently not yet available for this area, further comparison on the level 2 class was not performed.

Method 4: Exploration of temporal shifts of grassland types in the EUGW timeseries

Land cover trajectory analysis was carried out in some of the test sites to explore the major trends in grassland type changes in the period of 2016-2023. In this workflow, unique trajectories were created for each different pixel sequence, allowing for the quantification of persistent versus change pixels and identifying the trajectories responsible for the largest change (see A2_Table 5).

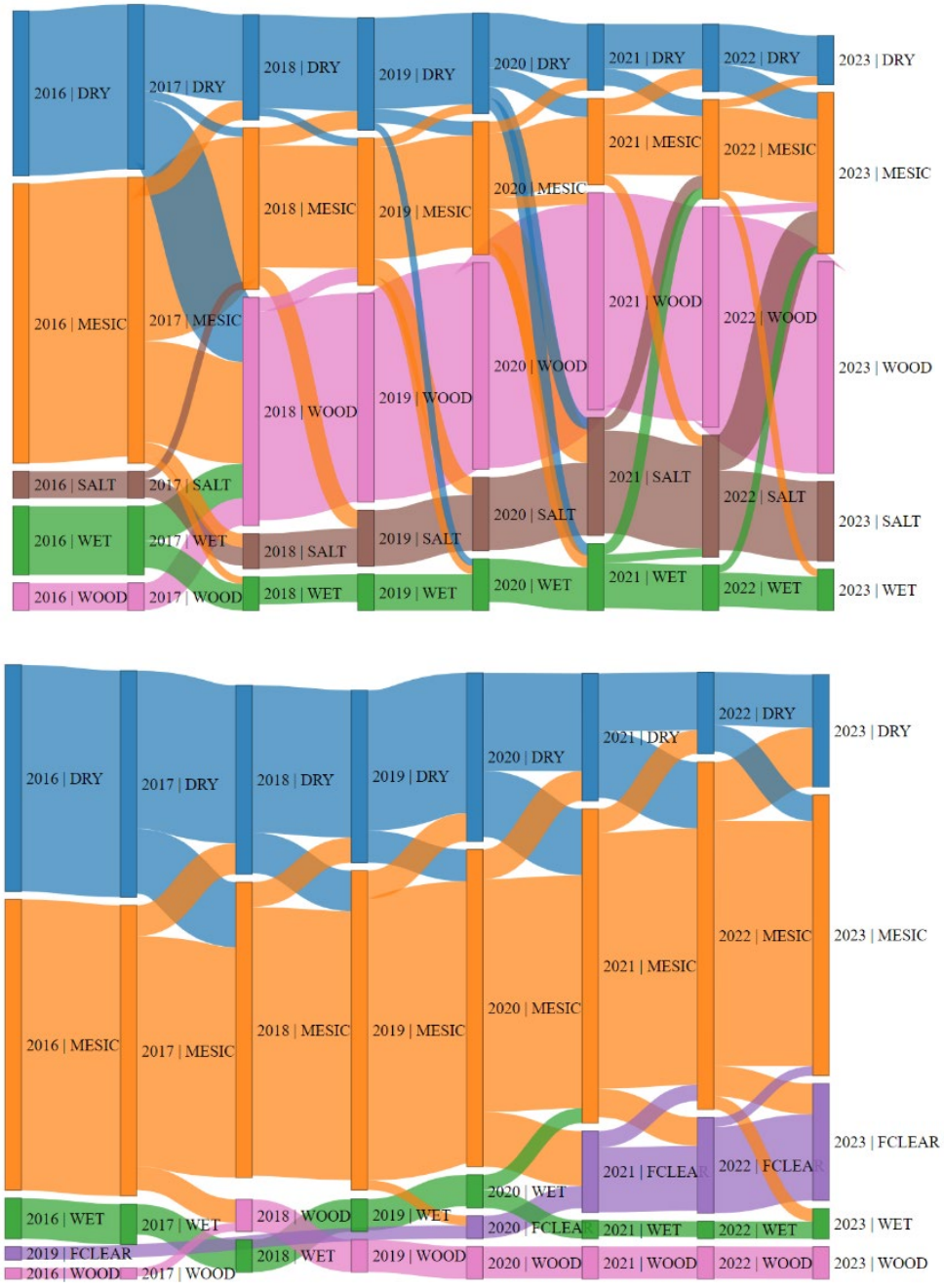
A2_Table 5 Area and temporal shift statistics for the two / three EUGW N2000 areas and the habitat change timeseries in them.

SAC/EUGW block	Total area analyzed (ha)	Persistent area (ha)	Change area (ha)	% persistent	% change
BE_CON_CON_61871	8511	7111	1400	83.5%	16.5%
DK_CON_CON_68746	5584	3526	2058	63,2%	36,8%
CZ_CON_CON_75111	3524	1820	1704	51,7%	48,3%

While the Belgian test site showed no extreme shifts in grassland typology across individual years, the Czech site, Milovice–Mladá, exhibited a marked degree of temporal shifts during the 2016–2023 time period. Only 51.7 % of pixels remained in the same habitat category throughout the entire period. The remaining pixels underwent varying levels of change: approximately 26.1% changed category once, 12% twice, and 4.6% three times. A smaller yet non-negligible proportion experienced four or more transitions. These values indicate substantial interannual shifts in the modelled grassland type classification, which are unlikely to reflect real in-situ changes in vegetation. The number of type change events in the Danish test site falls between the two other test sites, showing an intermediate volume of habitat shifts, with 63.2% of sites remaining in the same category.

These shifts can be visualized using a Sankey diagram (A2_Fig 25), where only larger pixel groups (>100 pixels within a group per year) and stronger flows between categories (>200 pixels participating in the transition between given typology levels) are visualized. The most pronounced shifts in the Milovice area occur between 2017 and 2018, when a substantial number of pixels shift into the category Sparsely wooded grassland (EUNIS code 27) from several other classes, most notably Dry grassland (21) and Wet and seasonally wet grassland (23). This transition corresponds to an interannual increase of more than 500 hectares of the sparsely wooded grassland category. Notably, this major reclassification is not accompanied by any visible change in vegetation structure when comparing available orthophotos from 2017 and 2019. This suggests that the observed shift is not driven by actual habitat dynamics, but rather by model-induced variability.

Appendix 3. EU Grassland watch validation



A2_Fig 25 Sankey diagrams representing pixel transitions in Milovice, Czechia (top) and Semois, Belgium (below) sites. Each vertical block represents one year, subdivided into the habitat classes assigned by the EUGW typology. Only pixel groups larger than 100 cells and flows exceeding 200 cells are displayed.

Conclusions and recommendations

The aim of this subtask of the Habitat Pilot was twofold: on the one hand, we wanted to compare the accuracy of the interim EUGW Typology and Land Cover component in different study areas throughout Europe. On the other hand, we wanted to test and roll out several complementary validation techniques to learn whether these could provide different insights into the model's accuracy and potential for improvements. Firstly, we applied indirect validation using ortho-imagery. Although this kind of validation approach is straightforward due to the relative scarcity of accurate in-situ data, it is limited in its validation potential for the Typology component. Nevertheless, in our experience, it has proven helpful in identifying certain potentially problematic areas by applying it to the overall grassland mask. Theoretically, the method can be harmonized to a great extent, using, e.g. sampling protocols and decision rules. However, in practice, validator bias was inevitable due to varying levels of a priori expert knowledge and equivocal 'grassland' definitions. As a result, in some areas, significant amounts of 'uncertain' validations were applied.

Regarding grassland detection, the overall accuracy was determined to be high, with a mean F1-score of 0.88 derived from thirteen distinct Special Area of Conservation (SAC) sites (ranging from 0.69 to 0.96). Lower accuracies were associated with two primary factors:

-The intrinsic ecological complexity of specific grassland sites (for example, Stora Alvaret in Sweden).

-The ambiguous interpretations of the sparsely wooded grassland class and forest clearing/fringes class detections.

It can further be argued that dealing with certain EUNIS level 2 classes (for example, wooded pastures, forest clearings and fringes, salt steppes, …) is most likely increasing the complexity of the underlying EUGW model architecture. However, from the perspective of grassland biodiversity conservation, these particular classes have a lower priority or, in some cases, are not even formally classified as grassland by ecologists. Consequently, it can be argued to omit these specific classes from future developments of the product. Moreover, it is pertinent to note that some of these classifications currently lack unequivocal in situ definitions across the European Union, which precludes adequate validation and accurate spatial modelling.

In our second approach, we illustrated that using local reference datasets is also beneficial for continued training and validating the level 2 classifications of the EUGW Type layer. From the test case in Belgium, observations from the previous approach were re-encountered, but additional areas of confusion were identified (e.g. wet grassland vs non-grassland wetland detections). Again, the issue of unharmonized definitions of EUNIS level 2 classes accounted partly for the loss of accuracy. Similar type and site-specific insights were gained from comparing vegetation maps from Milovice, Czechia and Parco Naturale dello Sciliar-Catinaccio in Italy. Cross-referencing RS maps with local reference maps proved to be a very valuable approach for the validation of the type level. Still, it is hampered by the ad hoc availability and temporal mismatch of these kinds of datasets.

The third and theoretically most accurate validation approach consisted of in-situ fieldwork to validate one or more specific EUNIS level 2 classes from the EUGW Typology product. Results varied across different areas and across different grassland classes, but generally revealed rather low accuracy of EUNIS level 2 classification in the tested areas (Denmark, Sweden, Czechia). It should be noticed

that some of these test sites were also amongst the areas with the lowest F1 scores in the grassland vs non-grassland validation approach, indicating specific grassland complexity at the site level.

Finally, some exploration of the temporal aspect of EUGW typology data was done using straightforward trajectory analysis to identify sudden shifts in the grassland class distribution throughout time. This approach again revealed high variance in confidence of the detected trends between sites. Abrupt shifts in grassland classes are very likely associated with model settings and varying quality of the input data rather than reflecting actual in-situ changes. Also, this approach illustrated high potential for detecting temporal anomalies and identifying aspects for future improvement.

Overall, we conclude from this site-level testing in the Habitat Pilot that the EUGW Type product in its current (standalone) form is lacking the crucial differentiation between relatively low-productive, species-rich grasslands and more intensively used, species-poor, agricultural grasslands. For example, it is unclear to what extent grass cropland (*Lolium* fields plowed regularly, green manure crops such as clover fields, etc.) is considered as part of the grassland mask and is subsequently included in the grassland monitoring statistics. This differentiation is indeed intended by the EUGW project to be achieved by overlaying with other features in the product suite, such as the Management and Productivity component. This seems a promising approach since direct grassland type detections using RS data are currently and will remain a challenge.

To what extent the EUGW Typology product in its current standalone form is useful for biodiversity monitoring purposes is also a question of scale. Accuracy of the typology on the site level proved to be very variable between classes and sites, so its usefulness for monitoring purposes on the site level remains limited. It should, however, be noted that we currently don't have a good estimate on the accuracy of the grassland types on the larger regional, national or transnational scale. To cover this, we recommend that a specific transnational validation of the EUGW products should be carried out in the future.

With our test cases in the habitat pilot, we clearly illustrate a high potential for accuracy optimization using detailed local reference data (in situ and existing data layers). Significant gain in accuracy and effectiveness for grassland biodiversity monitoring can be expected from unambiguously defining certain classes (for example, sparsely wooded pastures) or even reconsidering the actual relevancy of a class as a grassland component (for example, forest clearings).

In conclusion, satellite data products for monitoring grassland biodiversity should be viewed as complementary to—rather than a complete alternative for—current and other novel monitoring techniques. Within this complementary framework, metrics pertaining to management events, productivity, and detected changes will prove more efficient than focusing on straightforward type mapping alone.

References

Chytrý, M., Tichý, L., Hennekens, S. M., Knollová, I., Janssen, J. A. M., Rodwell, J. S., Peterka, T., Marcenò, C., Landucci, F., Danihelka, J., Hájek, M., Dengler, J., Novák, P., Zukal, D., Jiménez-Alfaro, B., Mucina, L., Abdulhak, S., Aćić, S., Agrillo, E., Attorre, F., Bergmeier, E., Biurrun, I., Boch, S., Bölöni, J., Bonari, G., Braslavskaya, T., Bruelheide, H., Campos, J. A., Čarni, A., Casella, L., Ćuk, M., Ćušterevska, R., De Bie, E., Delbosc, P., Demina, O., Didukh, Y., Dítě, D., Dziuba, T., Ewald, J., Gavilán, R. G., Gégout, J.-C., Giusso del Galdo, G. P., Golub, V., Goncharova, N., Goral, F., Graf, U., Indreica, A., Isermann, M., Jandt, U., Jansen, F., Jansen, J., Jašková, A., Jiroušek, M., Kącki, Z., Kalníková, V., Kavgacı, A., Khanina, L., Yu. Korolyuk, A., Kozhevnikova, M., Kuzemko, A., Küzmič, F., Kuznetsov, O. L., Laiviņš, M., Lavrinenko, I., Lavrinenko, O., Lebedeva, M., Lososová, Z., Lysenko, T., Maciejewski, L., Mardari, C., Marinšek, A., Napreenko, M. G., Onyshchenko, V., Pérez-Haase, A., Pielech, R., Prokhorov, V., Rašomavičius, V., Rodríguez Rojo, M. P., Rūsiņa, S., Schrautzer, J., Šibík, J., Šilc, U., Škvorc, Ž., Smagin, V. A., Stančić, Z., Stanisci, A., Tikhonova, E., Tonteri, T., Uogintas, D., Valachovič, M., Vassilev, K., Vynokurov, D., Willner, W., Yamalov, S., Evans, D., Palitzsch Lund, M., Spyropoulou, R., Tryfon, E. & Schaminée, J. H. J. 2020. EUNIS Habitat Classification: Expert system, characteristic species combinations and distribution maps of European habitats. Applied Vegetation Science 23/4: 648-675.

Chytrý M., Kučera T., Kočí M., Grulich V. & Lustyk P. (eds) (2010): Katalog biotopů České republiky. Ed. 2. Agentura ochrany přírody a krajiny ČR, Praha.

Schaminée J.H.J., Chytrý M., Hennekens S.M., Janssen J.A.M., Knollová I., Rodwell J.S. & Tichý L., 2018. Updated crosswalk of the revised EUNIS Habitat Classification with the European vegetation classification and indicator species for the EUNIS grassland, shrubland and forest types.Wageningen Environmental Research (WENR)

Appendix 4. Mapping of open grassland habitats

Partners involved:

- Pekka Hurskainen pekka.hurskainen@syke.fi (Finnish Environment Institute)
- Risto K. Heikkinen risto.heikkinen@syke.fi (Finnish Environment Institute)
- Mikko Kuussaari mikko.kuussaari@syke.fi (Finnish Environment Institute)

Background

Semi-natural grasslands and woodland pastures are habitat types that are shaped by traditional livestock grazing and mowing practices. Such practices created specific and diverse species assemblages which used to occur much more widely in the farmland landscapes of Europe, including Finland, than currently. Because of the cessation and absence of traditional management practices, these habitat types have become threatened, caused by pressures such as overgrowth and competing land use. Only 1 % of semi-natural grasslands in Finland remain, compared to the baseline extent in the 1950s. Thus, semi-natural grasslands are nationally assessed as the most threatened habitat group in Finland, with nearly all associated habitat types being critically endangered (Raunio & Kontula 2018). Without management, the extent and condition of these habitat types will further decline and may eventually go extinct in Finland.

Spatially explicit and up-to-date information on the extent of these habitat types is one of the key prerequisites for their regional and national conservation state assessments but such information still remains incomplete. A nation-wide survey of semi-natural grasslands and other traditionally managed agricultural biotopes was conducted during 1990's (summarized by Vainio et al. 2001). The survey was conducted mainly by inventories carried out by economic development centres (ELY, regional state administration) and Park & Wildlife Finland (Metsähallitus, Luontopalvelut). After the first nation-wide survey complementary inventories have been carried out in all parts of the country, and an updated summary on the state of the Finnish traditional rural biotopes was published last year (Forss 2024). The use of remote sensing in these surveys and mapping exercises has been limited only to on-screen orthophoto interpretation, and to our knowledge, no systematic efforts to include Earth Observation (EO) methods to this end have been made.

Apart from mapping the extent of traditionally managed semi-natural grasslands, there is a need to map also other open grassland habitat types with high nature values and may show potential to be restored towards a more species-rich semi-natural state through appropriate management measures. These include particularly long-term environmental fallows and arable pastures.

Objectives

The objective of this sub-task was to explore the possibilities to develop a semi-automatic mapping method for the open grassland habitats. Key preconditions for the methodological development were that the model should be based on the already existing and available pan-European and national EO and geospatial datasets, but also additionally on the in situ observations of different grassland habitats critical for model calibration and validation. An important methodological step forward would be developing approaches supported by EO data and other GIS data to sufficiently discriminate

between high-nature value, or biodiversity-rich, grassland habitats and other open vegetated man-made habitats.

In this subtask, we defined "**Open grassland habitats**" to *include* the following EUNIS habitat types:

- Mesic semi-natural grasslands (R2)
- Sparsely wooded grasslands/grazed woodlands (R7)
- Environmental fallows (V15, V13)
- Arable pastures (V11)

These agricultural/cultivated/grassland types were *excluded* from the subsequent analysis:

- *All other types* of arable land and market gardens (V1), including intensively cultivated grassland for forage and seed
- Cultivated areas of gardens and parks (V2)
- Urban parks, Golf courses (V3)
- Dry grasslands (R1)
- Seasonally wet and wet grasslands (R3)
- Forest fringes/edges and clearings (R5)

Rather than attempting to map each EUNIS type separately, the proposed broader grouping was seen to be more feasible to achieve sufficiently accurate results. This is because of limitations of in-situ data availability and quality on one hand, and spectral similarity of grassland types which hampers their discrimination potential from EO data on the other hand.

In addition to EUNIS classes R2 and R7, environmental fallows and arable pastures were included to the target class as these have the highest potential to be restored to a more species-rich semi-natural grassland state. The reason for excluding EUNIS R1 is its much smaller extent in Finland than R2 (Kontula & Raunio 2018), and consequently there is not enough in situ data for these. In addition, dry grassland is often found in small patches mixed within a mosaic grassland habitat. EUNIS category R3 was excluded because it is generally a less species-rich habitat type of modest conservation importance and with largely different plant and animal communities compared to mesic semi-natural grasslands. R5 was also excluded from this study, although they could have potential for restoration to semi-natural grasslands (Pykälä 2004, Jonason et al. 2014, Viljur & Teder 2016), if appropriately managed. Normally habitats included in R5 (typically clear-cuts) remain open only for a few years before their grass-featured vegetation becomes overgrown by planted trees.

In summary, this sub-task attempts to create a rough mask of **potential species-rich semi-natural grasslands**, rather than accurately delineate habitat type boundaries. The main use of this potential semi-natural grassland mask is to help Finnish habitat type experts in ELY centres and Wildlife and Parks Finland (Metsähallitus Luontopalvelut) to narrow down the search for grasslands and concentrate field work efforts. Another potential use is in national monitoring of the area of species-rich semi-natural grasslands in Finland, which has been requested for example in the national pollinator strategy (Ympäristöministeriö 2022). The methodological testing conducted in this subtask can also support the development of methods for mapping the grassland habitat of conservation concern in other European countries. Additionally, it can pave the way for EO and GIS data based monitoring of spatial changes in the occurrence of these habitats.

Recommendations

Based on the results and lessons learnt on the mapping and modelling of the open grassland habitats in this subtask, the following recommendations can be made for further developments and upscaling:

1. Importance of high-quality data for model calibration and validation, preferably collected *in situ* or sampled from trustworthy reference geospatial datasets. This is the most important part of building any classification model. It does not matter how many and how sophisticated EO / geospatial features are included, if the data for calibration (and validation) are not carefully curated, representative and extensive.The number of reference points for classes should be balanced relative to their actual proportions in the landscape.
2. Concurrent use of EO and other types of geospatial data. Information on land cover can be inferred from EO data, while information on land use and management typically requires other supportive information from geospatial datasets, such as Land Parcel Information System (LPIS) and Topographic map databases, or field inventories.
3. In the biogeographical regions that frequently experience cloud cover during the growing season, as well as low illumination and snow cover during spring and autumn, it can be difficult to acquire clear-sky satellite imagery. A mitigation strategy to address this is aggregating multiple images to monthly and annual composites, which can provide much more robust features. Even then, practice has shown that a single year of Sentinel-2 acquisitions is not enough to produce gap-free mosaics, but up to three years of data may be needed. If information is strictly needed from a particular month, Sentinel-1 monthly aggregates can offer an alternative as these are unaffected by atmospheric conditions like clouds, although they are much more sensitive to moisture conditions.
4. Monthly seasonal Normalized Differential Vegetation Index (NDVI) features, particularly from April, May and August, as well as NDVI annual amplitude, were found particularly useful for classifying open grassland habitats.
5. For further developments, it would be important to discriminate between short-term grasslands and long-term (perennial) grasslands.
6. Equally important would be the ability to better discriminate between intensively cultivated grasslands (for forage and seed production) and semi-natural grasslands. Currently, these are often grouped together in the same 'open grassland habitats' class.

Work description

Methods and Data

Summary of the mapping method. In this subtask, the method used to map open grassland habitats is the supervised image classification. Point samples of habitat and land cover types were used for calibration and validation of the model. As predictor variables, various features calculated from EO datasets were used. As classification method, non-parametric Random Forest (RF), including steps for recursive feature elimination and post-processing, was chosen based on existing knowledge of its good performance, easy parametrization and robustness in habitat classification.

Software and hardware requirements. The workflow was implemented with open-source tools. These included QGIS (visualization), GDAL (EO data processing), SAGA (computation of SAGA Wetness index), and R (data processing and analysis). Hardware consisted of a high-performance

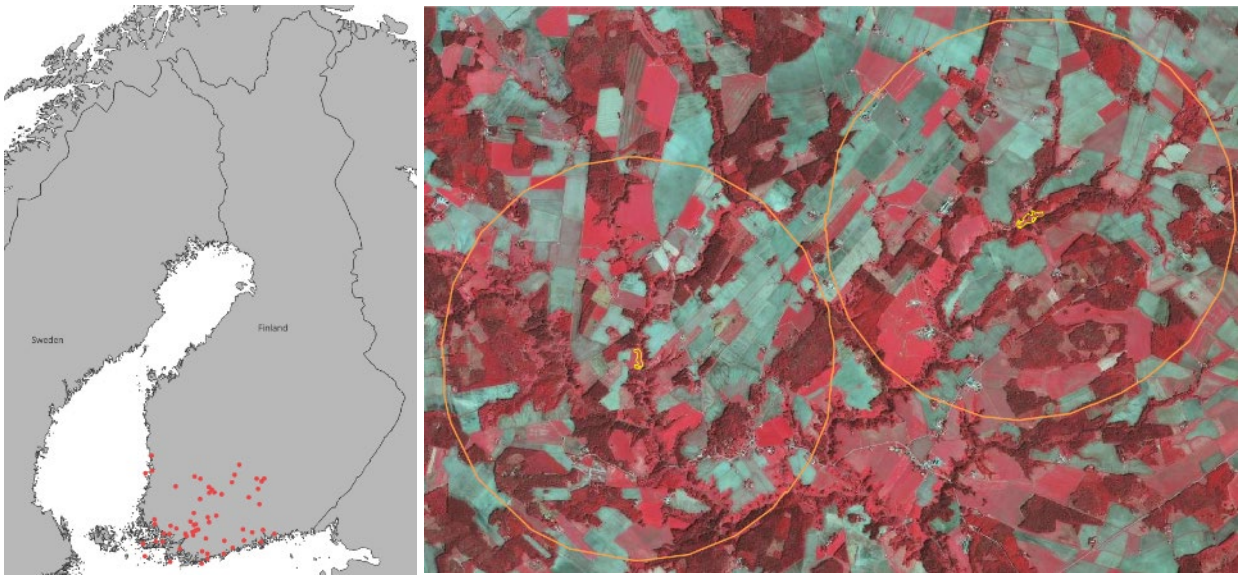
laptop with 64 GB RAM, Intel i9-12900H CPU with 20 cores available for parallel processing. Garmin GPSMap 65s was used to collect field validation data.

In situ data for open grassland habitats. The most important data for any classification task is sufficient, accurate and representative data for calibration and validation of the model. Since the method developed in this sub-task was implemented only in Finland, polygon data on biotopes from SAKTI geospatial database of protected areas, maintained by Parks & Wildlife Finland ('Metsähallitus Luontopalvelut'; Forss 2024), was used. First, a database search for mesic grasslands with a minimum extent of 0.5 ha across South-Western Finland was done. This returned a total of 972 biotope polygons (sites). From these, 30 Natura 2000-designated sites and 30 sites outside Natura 2000 were chosen by the experts on the grassland habitats and their biota. The criteria used for selecting the sites included field-assessed biodiversity value (Kemppainen 2017, Forss 2024), homogeneity and compactness (assessed against recent aerial photography and land cover data), condition (overgrown sites, sites with visible man-made structures were excluded) and geographical spread (distance between two selected sites should be at least 4 km).

Defining and processing the study area. For each of the 60 selected mesic grassland polygons, a buffer of 2 km radius was created. This resulted in 60 separate circular buffers between 1300 and 1600 ha in size, with a total extent of 57 066 ha (A4_Fig 1). To simplify the classification task, we used existing EO and geospatial data to constrain the classification model to look only in those land areas which had any potential of including open grassland habitat. In other words, these land cover types were excluded:

- Buildings, roads, other impervious surfaces and water, from High Resolution (HR) Land Cover Map (link), Copernicus HR Imperviousness Density (link)
- Vegetation (trees) taller than 2 m, from Canopy Height Model derived from national airborne LiDAR data (2019-2024, link), from some parts older LiDAR data (2008-2019, link) was used to get full coverage.
- All the above datasets were resampled to the same grid with 10 m pixel size to match with Sentinel 1 and 2 data. Then, pixels covered by the above-mentioned types were removed from the study area. For imperviousness, a density threshold of 20% was used.
- The side-effect of excluding non-focal pixels is that it tends to create small, isolated patches (clumps) of contiguous pixels surrounded by NULL values. In an additional processing step, area of all clumps was calculated and clumps smaller than 0.1 ha were removed. This 0.1 ha was thus used as a *Minimum Mapping Unit (MMU)* throughout the classification and post-processing.
- Another side-effect of excluding pixels was that many of the originally selected 60 grassland polygons were split to smaller polygons, totalling 130 polygons.
- After these steps, the remaining area to be classified was 26 453 ha.

Appendix 4. Mapping of open grassland habitats



A4_Fig 1 Location of the study area (left) and two selected mesic grassland polygons with their 2 km buffers (right).

Land cover and habitat types. To map open grassland habitats in the remaining landscape after exclusion of impervious, water and tall vegetation, six additional land cover / habitat types were defined to the classification legend. For these, no existing in situ data was available. Instead, existing national geospatial datasets were used as reference, from which random sample points for model calibration and validation were pooled (A4_Table 1 Classification legend and reference data.). As all of these were nation-wide datasets, they were cropped to the same extent of the study area and converted to polygons if they were in a raster data model.

A4_Table 1 Classification legend and reference data.

	Eunis type	**Reference data and metadata**
Open grassland habitats	R2, R7, V15, V13, V11	SAKTI (2024), link
Cropland	All other types of V1, V2	Land Parcel Information System (2024), link
Urban parks and Golf courses	V3	Topographic Map Database (2024), link
Forest clearings	R5	National HR Corine Land Cover (2024, *in production*) Masks of forest extent (2024), *used only in post-processing*, link
Bare rock	n/a	National HR Corine Land Cover (2024, *in production*)
Macrophyte communities	n/a	same as above
Peatbogs	n/a	same as above

Stratified random point sampling. The available reference polygons were used as a 'spatial study surface layer" from which sample points for calibration and validation were randomly drawn. However, as the occurrences of seven land cover classes (A4_Table 1 Classification legend and reference data.) were not evenly distributed in the study area, i.e. the 60 circular study sites, a proportional sampling approach was used:

- Calculate the extent and proportion of each class in the study area
- Remove polygons smaller than the MMU (0.1 ha)
- 1st iteration: for the most common classes, use 10 m negative buffer to avoid edge effects (e.g. shadow effects from adjacent tall trees), for others use the entire polygon extent (A4_Fig 2), and create 20 random points in each
- 2nd iteration: for each polygon, search and remove points which are too close to each other (30 m) to mitigate spatial autocorrelation
- 3rd iteration: based on the local proportions, randomly remove points from the most common classes (cropland, forest clearings, bare rock) and oversample (create more random points) for rare classes (urban parks and golf courses).

A4_Fig 2 Illustration of the stratified point sampling. Left: two reference polygons with edges (10 m) removed, middle: random points, right: remaining points after removing too close ones.

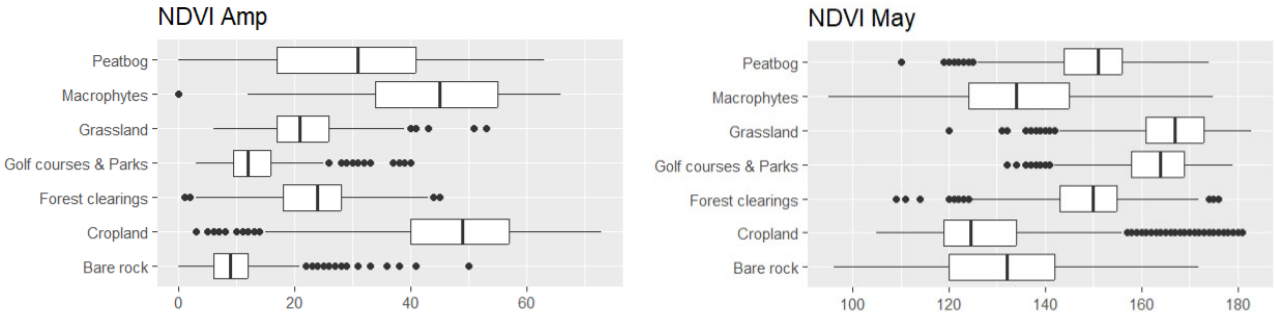
Additional in situ data collection. In June 2025, a limited field data collection campaign was done in nine of the 60 study sites. A total of 172 habitat type evaluation points were collected with the aim to validate the EO based habitat classification in the field. The main focus was to check the areas which were classified as open grassland in the RF model outputs and to collect more detailed information on their biodiversity value. Biodiversity value was assessed by surveying the plant community composition and especially the occurrence of positive and negative grassland indicator plants (Pykälä 2001). Plant species richness was estimated in six classes from very poor (only 1-2 vascular plant species) to very diverse (dozens of vascular plant species). Largely based on the occurrences of indicator plant species and plant richness, the sites were classified into the following grassland habitat classes: annual cultivated hay fields (for fodder), young (1-2-years old) environmental fallow, 3-5-years old environmental fallow, old (>5 years) environmental fallow, unmanaged abandoned grassland, recently (but not actively in the present year) managed semi-natural grassland and actively managed semi-natural grassland.

Calculate features for classification and build the data cube. Both pan-European and national EO data were employed to calculate explanatory features for the RF classification model (A4_Table 2). These included data from Copernicus and national data from Syke. The Sentinel-2 mosaic of 2024 was used as a reference, into which other raster layers were resampled. The output was one datacube (GeoTIFF) with a total of 60 features (layers). Lastly, the calibration and validation (cal/val) points were overlaid with the datacube, and feature values were extracted for each point.

A4_Table 2 EO datasets and derived features used as inputs for the Random Forest model.

Dataset, year	Features	Native spatial resolution	Source, metadata
Sentinel-2 best-pixel yearly mosaic (2024)	bands 2-8; 11-12, Transformed Chlorophyll Absorbtion Ratio (TCARI), Normalized Difference Moisture Index (NDMI)	10 m	Syke, link
Sentinel-2 NDVI annual statistical seasonal features (2024)	NDVI amplitude, maximum, minimum, mean, 10^{th}, 25^{th}, 50^{th} quartiles, sum	10 m	Syke, link
Sentinel-2 gapless mosaics of NDVI monthly maximum (2024)	Monthly NDVI maximum for April, May, June, July, August, September	10 m	Syke, link
Sentinel-1 Level 3 monthly weighted sum mosaics (2024)	VH + VV polarizations and ratio VH/VV for April, May, June, July, August, September	20 m	Copernicus Data Space Ecosystem, link
Sentinel-2 HR Vegetation Phenology and Productivity (HRVPP) (2024)	Start + End of season date, Season length, Season Max + Min Plant Phenology Index value, Season amplitude, Slope of green-up + green down period, Seasonal + total productivity	10 m	Copernicus Land Monitoring Service, link
Digital Elevation Model (DEM) (2019)	Elevation, Slope, Aspect, SAGA Topographic Wetness Index	10 m	National Land Survey + Syke, link
Canopy Height Model, National Airborne Laser Scanning (2009-2024)	Terrain Ruggedness Index, aggregated to 10 m using maximum and standard deviation	2 m	Syke, link1, link2

Data exploration. With feature values extracted for each cal/val point, summary statistics were calculated to see their distribution. Of specific interest was to see if there were points with missing values in any of the features, which then had to be removed for training of the RF model. In addition, boxplots were prepared for each feature, illustrating the distribution, spread and skewness of feature values for each class (A4_Fig 3). These were helpful visual tools to get a first impression of which features might be useful in classification, and where the potential outliers might occur.



A4_Fig 3 Example boxplots of NDVI annual amplitude and NDVI May maximum vs. target classes. Some classes have noticeably different spread than others, which is a good indication of a useful classification feature at least for those classes. Many potential outliers are also visible.

The Random Forest classification model. From the reference data points, 80 % was split for calibration (training of the model), and 20 % was set aside for validation (assessing model accuracy with data not used in model building). Proportions of the classes were kept the same in the model calibration and validation point sets. In the data splitting, the reference polygons were used, not points. This was to ensure that calibration and validation points were not originating from the same polygon.

Feature selection is an important step in building a machine learning classification model. Feature selection algorithms attempt to remove unnecessary features which do not improve classification accuracy (but do not necessarily remove correlated features). The expected outcome of this is a more simplified, parsimonious model which also enhances model interpretability. Here, a Recursive Feature Elimination (RFE) algorithm was implemented for RF, which uses external resampling and 10-fold cross-validation. RFE fits the model first to all features and ranks each based on its importance to the model. At each iteration of RFE, the worst performing feature is removed. Then the model is refitted and performance reassessed, and iterated again, until no features are left.

After running the RFE algorithm, it was discovered that model overall accuracy did not improve after 24 most important features, thus 36 features were discarded. Feature selection was run only once, but the remaining parts of model building – training, prediction, accuracy assessment and validation – were iterated multiple times, as training data was modified (see later for more).

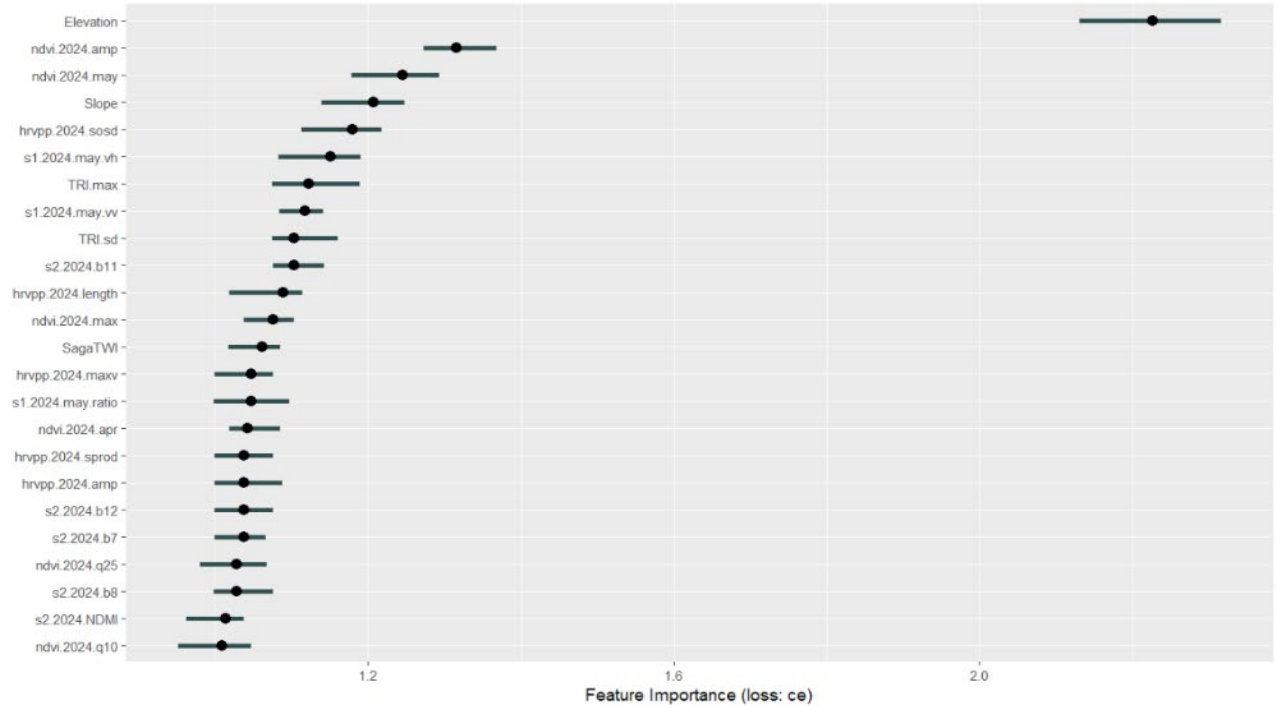
The RF model was then trained with the selected variables and calibration points. The trained model was then fitted (predicted) to 'novel terrain', i.e. to every pixel of the datacube and the study area. Each pixel was assigned into one of the seven habitat or land cover types based on the highest probability of belonging to that type.

In situ validation measures of the model. In-situ points collected in the field were added to the validation data set. The classification accuracy of the RF model was then assessed *quantitatively* with independent validation data by creating a confusion matrix. The following accuracy metrics were derived from the matrix: Overall accuracy with 95% confidence intervals and Kappa value for the whole model; Precision, Recall and F1-value for each class.

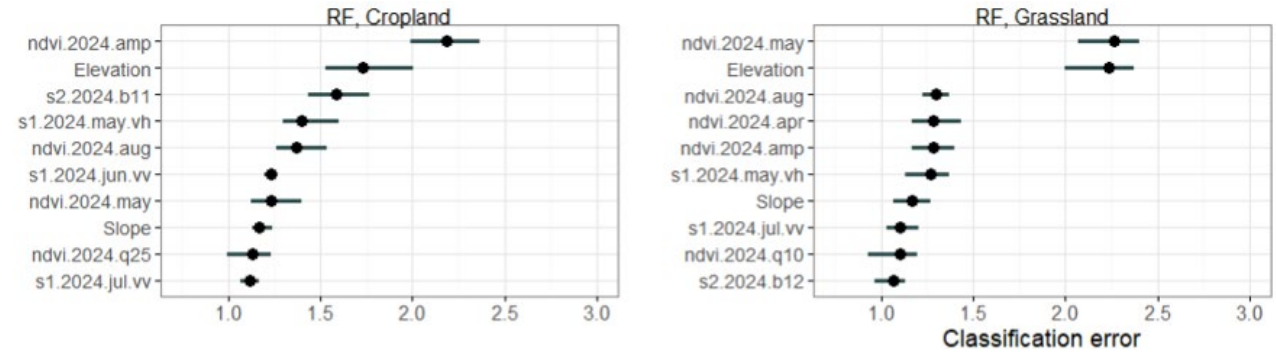
An important part of the validation is also to look behind the numbers. A "look-and-feel" *qualitative* visual comparison to other GIS/EO data can yield impressions on how realistically the classes are spatially distributed with respect to the existing knowledge. This was done in QGIS by overlaying and comparing the classification output raster to various available reference data, such as SAKTI mesic grassland polygons, the LPIS data, and recent aerial photographs and satellite data. This gave insights on which classes the model outputs tend to mix, and might thus need more attention in the next classification iteration round.

Feature importance and outlier removal. We focused particularly on investigating which features were most important for the whole model (A4_Fig 4), and more specifically, what were most important features for distinguishing grassland habitats from other habitat and land cover types (A4_Fig 5). Model class reliance (MCR) using classification error (ce) loss function was calculated for each class and feature. The MCR is a unitless metric, which describes the highest and lowest degree to which RF within a given class may rely on a specific feature for prediction accuracy. In practical terms, the MCR value (ce) of 2 can be interpreted as strong model reliance on the feature, while a value closer to 1 signifies less or no reliance.

Appendix 4. Mapping of open grassland habitats



A4_Fig 4 Feature importance rankings of the RF model.



A4_Fig 5 Ten most important features in the RF model for classifying cropland (left) and grassland (right).

To improve model performance, it was useful to inspect feature importance for those classes which were frequently confused with grassland habitats in the model outputs, and to check if there were outliers in the training data for those classes. For example, it was found that the model often confused open grassland habitats and forest clearings with cropland. In addition, NDVI amplitude was an important feature for all of these. According to the box plot of NDVI amplitude (A4_Fig 3), there were outliers in each of these classes. These outliers might be true observations with a significantly different NDVI amplitude than the other members in these classes, mixed pixels, or errors in the reference data from where the points were randomly sampled. Whatever the case, such observations can compromise the model's fit. Thus, it was decided that they should be removed using the following procedure: Inspect the boxplot, find the 1st and 3rd quartiles, calculate interquartile range (IQR), find lower (Q1 – 1.5*IQR) and upper (Q3 + 1.5*IQR) limits for outliers, and remove points identified as outliers. In total, 492 points (~7 %) were removed from cal/val points. After this

4th iteration round of cal/val points, the model was trained again with the reduced training points, predicted to the study area, and a new validation was done.

A4_Table 3 shows a summary of the extent and share of each habitat type or land cover class in the study area (based on existing reference data), sum of all polygons before and after filtering, number of random points available after each iteration round, and the final number of points available for the RF model training and validation.

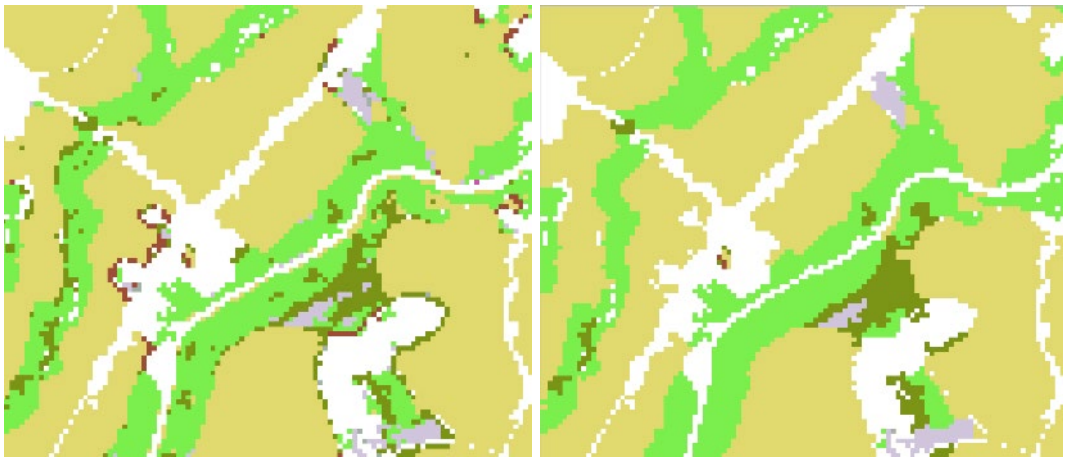
A4_Table 3 Statistics of the polygons and randomly sampled points from existing reference data.

	Cropland	Peatbogs	Macrophytes	Urban parks & Golf courses	Open grassland habitats	Forest clearings	Bare rock
Extent in study area (ha)	11283	144	271	168	117	799	186
Share of study area	41,02 %	0,52 %	0,98 %	0,61 %	0,43 %	2,91 %	0,67 %
Sum of all polygons	3454	268	1149	114	130	1262	1309
Sum of polygons after filtering with MMU	2855	108	245	54	73	449	234
Sum of random points (1st iteration)	56920	2160	4060	1080	1460	8300	4680
… after removing points too close (2nd iteration)	25168	558	731	294	564	2332	953
… after under- and oversampling (3rd iteration)	2960	558	731	555	564	1166	635
… after removing outliers (4th iteration)	2656	558	731	501	544	1052	635
Sum of calibration points (80 %)	2151	433	565	405	460	846	351
Sum of validation points (20 %)	505	98	133	86	84	206	95
Sum of validation points + in situ	549	98	133	86	210	208	95

Post-processing. The outputs of the qualitative and quantitative validation conducted in the previous steps were related to the accuracy of the RF model. However, the predicted outputs of the model are not necessarily the ultimate end result, as there are still ways to improve them through post-processing. Carefully designed post-processing can result in a much more accurate and useful final end-product.

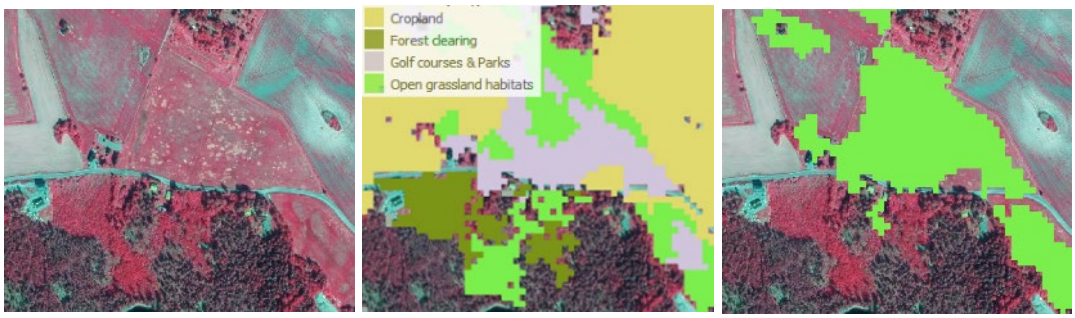
After several rounds of model building and in-situ data adjustment iterations, the results were deemed to be 'good enough'. Then, the following post-processing steps were followed:

1) **Step 1. Generalization to MMU**. Using a **sieve** filter, "noise" in the classification was reduced by changing small clumps of pixels (less than 0.1 ha in size) with a class that is different from the surrounding pixels, to the class of the largest neighbouring clump (A4_Fig 6).



A4_Fig 6 After running the sieve filter (right), clumps smaller than MMU (0.1 ha) are reclassified to their largest neighbor clump in the classification output (left).

2) **Step 2. Extraction of open grassland habitats.** All pixels classified by RF as open grassland habitats were extracted from the classification map, creating a grassland mask. The following post-processing steps were then run only for this mask.
3) **Step 3. Adjustments to urban parks and golf courses**. The model was successful in mapping the known golf courses and urban parks, but overpredicted their extent to actual open grassland habitats. Typically a single grassland patch was partially accurately identified, and partially classified to this very similar class. Since a very good reference data for urban parks and golf courses was available for our study, we made an assumption that their extent was already well known. Thus, all overpredicted golf course and urban park pixels outside the reference data were reclassified during the post-processing into open grassland habitats (Figure 7).
4) **Step 4. Corrections between forest clearings and grassland.** The RF model identified forest clearings, verified as true clearings from aerial photographs and land cover maps, with sufficient accuracy. However, the RF model overpredicted grassland to forest clearings. Forest clearings also occurred sometimes as mixed features inside grassland patches. To correct these misclassifications, forest masks from Finnish Forest Centre were used to limit the extent of forest clearings, and overpredicted pixels outside the reference data were reclassified into open grassland habitats (A4_Fig 7).



A4_Fig 7 Aerial photograph (left), RF classification after sieve (middle), grassland mask extracted from the classification, after adjustments made for overpredicted urban parks, golf courses, and grassland.

5) **Step 5**. Fill small holes. As the last post-processing step, holes inside the grassland mask (clumps of pixels entirely surrounded by grassland) were identified and their area calculated. Holes smaller than 0.02 ha (1-2 pixels) were then filled, i.e. reclassified to the grassland mask. As these holes inside grasslands were mostly a result from removing tall vegetation, this post-processing step helped to map the sparsely wooded grasslands in more detail.

Results

Model overall accuracy and confusion matrix. Quantitative assessment showed that the RF model reached overall accuracy of **87.9 %**, with additional values of 95 % CI [86.1 – 89.7], and kappa 0.844. However, after the post-processing step 1, many wrongly classified small clumps of pixels were generalized to the class of the largest neighbour. The overall effect of these reclassifications were positive, as model overall accuracy **after sieve** increased to **90.4 %**, with 95 % CI [88.8 – 91.9], and kappa 0.876. The confusion matrix of the RF model, without post-processing, is shown in A4_Fig 8. As a comparison, confusion matrix after the sieve is shown in A4_Fig 9. Looking at the latter figure, the post-processing step 1 helped considerably to reduce confusion between open grassland habitats, forest clearings, and golf courses & urban parks, but not entirely.

A4_Fig 8 Confusion matrix of the RF classification model (no post-processing).

Appendix 4. Mapping of open grassland habitats

A4_Fig 9 Confusion matrix of the RF classification model after sieve.

Class-wise model accuracy. Accuracy metrics precision, recall and F1 were also calculated per class, and these are shown in A4_Table 4. Based on the F1-value - which is the harmonic mean of precision and recall - cropland, forest clearings and bare rock were classified with highest accuracy. The F1-value for open grassland habitats was somehow more moderate (0.824), and also here a significant improvement from sieve post-processing was apparent (0.775).

A4_Table 4 Class-wise accuracy metrics of the RF model after sieve post-processing (step 1).

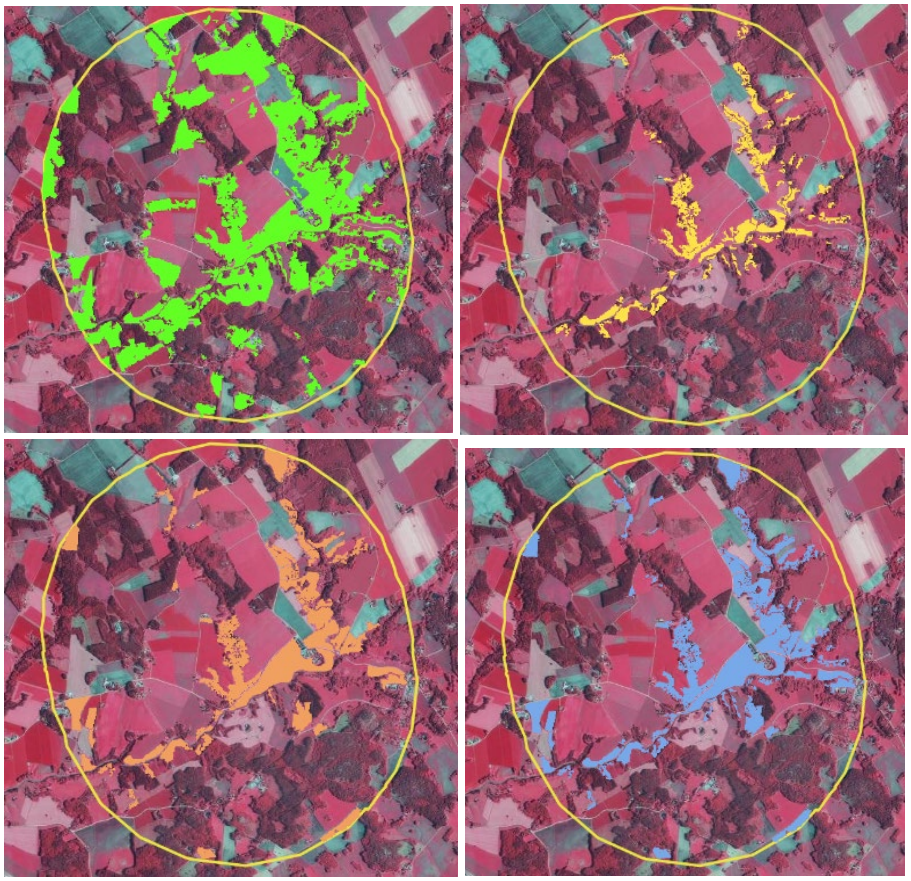
	Cropland	Peatbogs	Macrophytes	Urban parks & Golf courses	Open grassland habitats	Forest clearings	Bare rock
Precision	0.963	0.794	0.939	0.826	0.802	0.917	0.932
Recall	0.953	0.867	0.812	0.826	0.848	0.962	0.863
F1	0.958	0.829	0.871	0.826	0.824	0.939	0.896

Comparison to reference grassland masks. The accuracy metrics presented above refer to the accuracy of the RF model. Additional post-processing steps 2-5, described above, were done for the grassland mask, attempting to create a more useful product for mapping the potential open grassland habitats. In order to assess the usefulness of this product, three **reference grassland masks** were generated from the best available national data sources via the following data management steps:

- From the LPIS data, all types matching to the definition of “open grassland habitats” within the study area were chosen (2294 polygons)
- From the SAKTI data, all mesic grasslands within the study area were chosen (1389 polygons)
- These data were rasterized to same 10 m grid with the grassland mask (derived from RF model)
- Clumps of pixels smaller than MMU were removed
- LPIS and SAKTI masks were combined to generate a third “all-inclusive” mask

- Same procedure of excluding non-focal pixels was executed as was done for the study area (vegetation > 2 m, built-up)

After these processing steps, a pixelwise comparison was done between the grassland mask (hereafter referred as “product”) and the three reference masks: SAKTI + LPIS, and SAKTI / LPIS alone (A4_Fig 10).



A4_Fig 10 Examples of the grassland product in Rekijokilaakso, a Natura 2000 -site (green), and the three reference grassland masks and they polygons shown as different colours: SAKTI mesic grasslands (yellow), LPIS open grassland habitats (orange) and combined SAKTI+LPIS (blue).

First, comparison to the combined SAKTI+LPIS reference was done. Statistics were calculated for pixels where there was agreement between the product and reference, for pixels which were in the reference but not in the product, and vice versa (A4_Table 5). A total of 2320.5 ha (70.2%) of open grassland habitats were found in the product, which were also in the reference grassland mask (i.e. true positives), whereas 985.8 ha (29.8 %) were false negatives. In addition, a total of 5220.2 ha of “false positives” were found, but these might as well be new open grassland habitats that were previously unknown in the reference data.

A4_Table 5 Comparison between the grassland product and reference grassland mask, including all SAKTI and LPIS open grassland habitats.

Comparison outcome	Number of pixels	Extent (ha)	How big proportion of grassland in reference was found/not found?

True positive: grassland in both product and reference	232052	2320.5	70.2 %
False negative: grassland not in product, but in reference	98584	985.8	29.8 %
False positive: grassland in product, not in reference	522024	5220.2	n/a

Then, similar comparisons were done between the grassland product and SAKTI mesic grasslands (A4_Table 6) and LPIS open grassland habitats (A4_Table 7). This comparison gives more hints on which types of grasslands were more accurately mapped. A very promising result was that 88.0 % of mesic grasslands (505.4 ha) in the SAKTI was found in the product (true positives), and only 12.0 % (69.1 ha) were missed (false negatives). The mapping method was less successful in locating other types of open grassland habitats when compared to the LPIS data – just 69.6 % (2179.8 ha) were labelled as true positives, with 30.4 % (950.3 ha) false negatives.

A4_Table 6 Comparison between the grassland product and SAKTI reference grassland mask.

Comparison outcome	Number of pixels	Extent (ha)	How big proportion of grassland in reference was found/not found?
True positive: grassland in both product and reference	50539	505.4	88.0 %
False negative: grassland not in product, but in reference	6906	69.1	12.0 %

A4_Table 7 Comparison between the grassland product and LPIS reference grassland mask.

Comparison outcome	Number of pixels	Extent (ha)	How big proportion of grassland in reference was found/not found?
True positive: grassland in both product and reference	217980	2179.8	69.6 %
False negative: grassland not in product, but in reference	95027	950.3	30.4 %

Discussion

The mapping method developed and tested in this subtask works reasonably well. Mapping accuracy metrics (precision 0.802, recall 0.848, F1 0.824) are sufficient towards the objective of creating a rough mask of potential open grassland habitats. It was also demonstrated that post-processing the classification output with geospatial analysis methods can further improve the results to a more useful product. However, if exact habitat boundary delineations are needed, field work including vegetation surveys would often be a necessary additional step.

Comparison of the grassland product from this subtask to the national reference grassland mask (SAKTI database) showed that about 88 % of known mesic grasslands could be correctly identified. On the other hand, 5220 hectares of open grassland habitats, which were not in either of the reference grassland data sets, were found. It is not known if these are true or false positives. Clearly,

such an assessment requires a verification process conducted based on a sufficiently extensive field inventory.

Moreover, a further comparison to the grassland mask of EU Grassland Watch (EU GW) would have been a very useful comparison point for our national data focussed verification exercise. However, unfortunately the updated versions of EU GW products were not available for the grasslands in Finland during the project.

Important features for mapping open grassland habitats. Topographic features and features informing on seasonal/phenological patterns were most important for the whole RF model (Figure 4). Elevation as a single feature stands out from the rest with very high classification error value, which can be partially explained by the landscape characteristics: many of the 2-km buffers of the study area are situated next to the coastline, or at the archipelago of Baltic Sea, where macrophyte communities and bare rock habitats are common, while further inland they are much rarer. NDVI amplitude was ranked as the second most important variable in the RF models. NDVI was calculated as pixelwise annual maximum NDVI minus the 25-quantile NDVI of the current year and two previous years and thus characterizes the annual dynamics of green vegetation. NDVI maximum in May was the third most important, which in Southern Finland is typically the month when vegetation green-up begins. Many of the HRVPP parameters from Copernicus were useful features, as well as Sentinel-1 May composites, and terrain ruggedness features from Airborne laser scanning data. Single bands from the Sentinel-2 annual mosaic were of lesser importance, apart from band 11 (Shortwave infrared).

For open grassland habitats (A4_Fig 5), the most important feature was NDVI May. Elevation was nearly as important, which was surprising as the training data included mesic grasslands from very low elevation at the coast to higher elevations inland. NDVI maximum values in August and April, NDVI annual amplitude, and VH polarization of Sentinel-1 backscatter of May were also equally important features.

Maturity of the method. It is difficult to say how useful the final output is without proper testing and feedback from its intended users. Technically, the method is mature enough to be scaled up for the whole Southern Finland and this could potentially improve the mapping accuracy as more reference data for calibration of the model would become available. All EO and geospatial data used in this module are available nation-wide. On the other hand, processing and analysing the data would have to be installed into a high-performance cloud computing system, requiring more resources. More time would be needed also for curating the reference data, which can have errors due to various reasons.

Transferability of the method outside Finland. The method could be tested and transferred to other European countries, assuming comparable geospatial and EO datasets are available. All relevant data processing and analysis steps are ready for further applications. Most EU countries should have a LPIS database with annual crop type information, and existing point or polygon data of semi-natural grasslands as these are needed inputs for the model. Other semi-natural grassland types could easily be added to the classification legend. Reference data for the other land cover types should be available from national land cover maps and topographic databases.

On the EO data used, Sentinel-2 HRVPP parameters as well as Sentinel-1 monthly composites are analysis-ready data available across Europe. The Sentinel-2 mosaic was created with the Copernicus Sentinel-2 Global Mosaic-service (link), which is freely available to users. The annual

NDVI statistical seasonal features and NDVI monthly maximum products from Sentinel-2 are relatively straightforward to calculate and customize with available scripts in e.g. Sentinel Hub or OpenEO. Many countries have their own national HR Digital Elevation Models (DEM), and if not, the Copernicus DEM at 10 m resolution can be used (link). Canopy height models from Airborne Laser Scanning, as well as aerial photographs for visual verification, are EO datasets that have to be sourced nationally.

References

Forss, S. 2024. Perinnebiotooppien valtakunnallisen inventoinnin päivitys. Yhteenveto Suomen perinnebiotooppien tilasta 2023. Suomen ympäristökeskuksen raportteja 28/2024. 83 s. https://helda.helsinki.fi/items/f3f65550-8ad2-4036-9fff-94bbd404cdc4

Jonason D, Ibbe M, Milberg P, Tunér A, Westerberg L & Bergman KO 2014. Vegetation in clear-cuts depends on previous land use: a century-old grassland legacy. Ecol. Evol. 4, 4287–4295.

Kemppainen, R. 2017. Perinnemaisemien inventointiohje. Varsinais-Suomen elinkeino-, liikenne- ja ympäristökes-kus. Raportteja 25/2017. 89 s.

Kontula, T. & Raunio, A. (toim.). 2018. Suomen luontotyyppien uhanalaisuus 2018. Luontotyyppien punainen kirja – Osa 1: Tulokset ja arvioinnin perusteet. Suomen ympäristö 5/2018. Suomen ympäristökeskus ja ympäristöministeriö.

Pykälä, J. 2001. Perinteinen karjatalous luonnon monimuotoisuuden ylläpitäjänä. Suomen ympäristö 495. 205 s.

Pykälä, J., 2004. Immediate increase in plant species richness after clearcutting of boreal herb-rich forests. Appl. Veg. Sci. 7, 29–34.

Toivonen, M., Huusela, E., Hyvönen, T., Marjamäki, P., Järvinen, A. & Kuussaari, M. (2022) Effects of crop type and production method on arable biodiversity in boreal farmland. Agriculture, Ecosystems & Environment, 337, 108061.

Vainio, M., Kekäläinen, H., Alanen, A. & Pykälä, J. 2001. Suomen perinnebiotoopit. Perinnemaisemaprojektin valtakunnallinen loppuraportti. Suomen ympäristökeskus, Helsinki. Suomen ympäristö 527. 163 s.

Viljur, M.-L. & Teder, T. 2016. Butterflies take advantage of contemporary forestry: Clear-cuts as temporary grasslands. Forest Ecology and Management, 376, 118–125.

Ympäristöministeriö 2022: Kansallinen pölyttäjästrategia ja toimenpidesuunnitelma. Ympäristöministeriön julkaisuja 2022: 9. http://urn.fi/URN:ISBN:978-952-361-246-4

Appendix 5. Measuring vegetation height and cover with remote sensing tools

Partners involved:

- Jesper Erenskjold Moeslund jesper@ecos.au.dk (Aarhus University)
- Cuong Ngo nm.cuong@ecos.au.dk (Aarhus University)
- Stien Heremans, stien.heremans@inbo.be (Research Institute for Nature and Forest - INBO)
- Patrik Oosterlynck, patrik.oosterlynck@inbo.be (Research Institute for Nature and Forest - INBO)
- Hans Gardfjell, Hans.gardfjell@slu.se (Swedish University of Agricultural Sciences)
- Sven Adler, sven.adler@slu.se (Swedish University of Agricultural Sciences)

Background

In monitoring habitat conditions, knowledge of the distribution of and variation in vegetation height is important (Hunter *et al.* 2025). Also, being able to monitor how this develops over time provides central evidence on which path a given habitat is on, e.g. to which degree tall herb and shrub encroachment takes place or not (Xu *et al.* 2023). The vegetation height characteristics of a given habitat reflect a multitude of different important aspects of that habitat's condition, for example the grazing intensity, nutrient load, and plant diversity.

At present, vegetation height is typically measured in the field during field inventories. Hence, usually we only have point or plot-based information on this habitat feature. Since the lowest vegetation (down to at least 10 cm) is among the most important characteristics to be measured for certain habitats, especially grasslands, large-scale remote sensing approaches based on e.g. satellites are not the most suitable tools for providing an alternative to field measurements (Reinermann *et al.* 2020). However, advances in drone technology, notably drone-borne laser scanners are now at a promising stage where testing this technology's ability to provide reliable wall-to-wall vegetation height monitoring data for nature areas makes sense. In this subtask, we aimed to test one of the newest drone laser scanning methods to collect and produce wall-to-wall vegetation height maps suitable for habitat condition monitoring. Moreover, as part of this exercise, we aimed to evaluate how this method upscales, i.e., how useful and reliably and easily it is applied in other areas than where it was developed for.

Recommendations

Accurately monitoring vegetation height is crucial to understanding habitat condition in dynamic ecosystems. Parameters related to vegetation height provide valuable insights into ecological processes such as shrub encroachment, nutrient status, grazing pressure, and plant species diversity. Traditionally, this metric is collected through field surveys, limiting observations to plot-based samples and requiring significant manual effort. The work described in this subtask demonstrates that drone-borne LiDAR offers a promising way to scale up vegetation height monitoring across larger landscapes and diverse ecological contexts.

Our results provide encouraging evidence that vegetation height derived from drone point clouds can be used for habitat condition monitoring. Despite operating across different countries and environments, our tests showed that a deep learning model originally trained in one region (Denmark) was able to deliver comparable performance in another (Belgium), even when different drones and LiDAR sensors were used for data collection. This indicates a high level of scalability and transferability than we had expected. The ability to operationalize models across borders increases the potential to build harmonized monitoring systems at national and international levels.

However, our attempts also highlighted challenges that need further attention. The model was not always scalable (i.e., having the same accuracy as at the original site) and despite reasonable classification performance for medium vegetation, predicting low vegetation height (<25 cm) — often critical for grassland and wetland quality — proved difficult. Misclassifications between low and medium vegetation were common and likely reflect the complexity of working with inherently ambiguous ground truth. The validation of vegetation height, which is gradual rather than categorical, remains a critical methodological hurdle. Our approach of binning continuous height measurements into discrete classes masks some of this complexity. Therefore, advancing validation protocols that recognize the uncertainty in both the automated predictions and the field measurements is essential.

We recommend that further methodological refinement of the deep learning approaches used to process drone LiDAR data is undertaken. In particular, efforts should be made to improve model performance so it more closely reflects human-like estimation of vegetation height — especially in ecotones and areas with fine-scale variation of the vegetation. This could include further exploration of sensor fusion, additional model training on diverse datasets, and improved post-processing routines.

Once matured, these methods offer a compelling vision: truly wall-to-wall, regularly updated vegetation height data that can guide adaptive habitat management and provide early warnings of ecological change. By taking full advantage of drone technology and advances in deep learning, we can move closer to real-time monitoring of habitat condition at a level that meets the needs of conservation practitioners and policymakers alike.

Work description

Methods and data

We used data collected by three different drone-borne laser-scanners in four sites in Denmark, Sweden, and Belgium. The method that we tested was developed as part of the Horizon Europe project Modern Approaches to Monitoring of Biodiversity (MAMBO). While the development of this method is still to be finalized, i.e. it is not officially published due external reasons, it will be made publicly available as soon as possible. The method is based on deep learning, as deep learning models have been shown to be a suitable methodological choice for such data. This is particularly due to their ability to scale with a high number of parameters, corresponding to huge amounts and variation in the data. Furthermore, deep learning is necessary to handle both high-level context and fine-grained local detail.

Data collection

The test areas were Molslaboratoriet and Understed Bakker in Eastern and Northern Jutland respectively (Denmark), Örsten in Eastern Sweden, and Zammelbroek in the Flanders region of Belgium. All the point cloud data were collected using drones, see details in S5_Table 1.

S5_Table 1 Lidar sensor and data collection details

Location	Lidar sensor	Return no's per pulse	Year
Molslaboratoriet, Denmark	DJI L1	3	2023
Understed Bakker, Denmark	DJI L2	5	2024
Örsten, Sweden	DJI L1	3	2024
Zammelsbroek, Belgium	YellowScan Surveyor Ultra 2	3	2024

Drone borne lidar data was collected during the growth season in all countries and using the same flying settings. Flying height was 60 m asl and speed was 5 m/s. Data were captured with 70% forward overlap and 50% side overlap, yield point clouds with generally > 1000 points/m^2.

Data processing

The point cloud data from Denmark and Sweden (DJI equipment) are provided out-of-the-box as a colorized point cloud after processing the collected data in the DJI Terra software. The data from Belgium were partially colorized after processing, but some misalignment existed between the point cloud and the orthophotos recorded by the drone simultaneously, possibly due to slight GPS drift or uncorrected photogrammetric distortion. Because of that, we used the nearest neighbor algorithm to colorize the uncolored points with their nearest colorized neighbors in the Belgian point cloud.

In a subsequent step, we converted the three bands red, green and blue into the CIELAB color space (LAB) as the inputs for our deep learning models. CIELAB colors are more suitable for our purpose as they provide a more uniform color perception and separate lightness from color information. The CIELAB color space, often referred to as Lab*, is designed to approximate human vision. It expresses color with three values: L* for perceptual lightness, and a* and b* for the four unique colors of human vision: red, green, blue, and yellow. This contrasts with the RGB color space, which is based on the combination of red, green, and blue light and is more device-oriented[JM1] .

Next, we divided the point cloud into smaller regular tiles to fit the limited memory of the computer we used. For Molslaboratoriet, Örsten and Zammelbroek, we used 100mx100m tiles. For Understed Bakker, we had a higher density of points and hence used smaller 40mx40m tiles. In all the areas, we created tiles having 20-meter overlap to reduce edge effects[JM2].

Modelling

The core task here is *semantic segmentation* of a 3D point cloud. This means assigning a specific class label (like "ground", "vegetation", or "building") to every single point in the cloud. This is a dense prediction task, requiring both high-level context and fine-grained local detail. In the following we describe the MAMBO model in detail.

The algorithm combines well-known U-Net architecture with the specialized SparseConvNet library from Facebook Research. This approach is highly effective for 3D point cloud semantic segmentation because:

- *U-Net's encoder-decoder structure with skip connections* is excellent for learning both global context and local details, which are crucial for accurate segmentation.
- *SparseConvNet* provides the necessary tools to implement this U-Net efficiently on sparse 3D point cloud data, avoiding the computational and memory overhead of dense 3D convolutions. Point cloud data is by nature “sparse”, meaning that there are not necessarily points in all 3D quadrats (voxels) in the space one is processing. Processing all the empty data is costly and can be avoided using SparseCNNs.

This model takes raw 3D point coordinates and attributes like Return Number (the sequential position of a specific point within the sequence of returns from a single laser pulse) and Number of Returns (the total number of distinct echoes, or returns, that a single emitted laser pulse generates as it travels towards the ground). Then it processes them through a sparse U-Net to learn hierarchical representations derived from actual points, rather than empty space, and finally outputs a class prediction for each point, effectively performing semantic segmentation of the point cloud.

The model was trained on the point cloud data from Molslaboratoriet and from the Hessigheim 3D (H3D) benchmarking dataset[1]. The latter was collected in 2018 and 2019 in the village Hessigheim in Germany. The Hessigheim data was recorded using a drone with both a LiDAR scanner and cameras, providing both a dense point cloud and a 3D textured mesh.

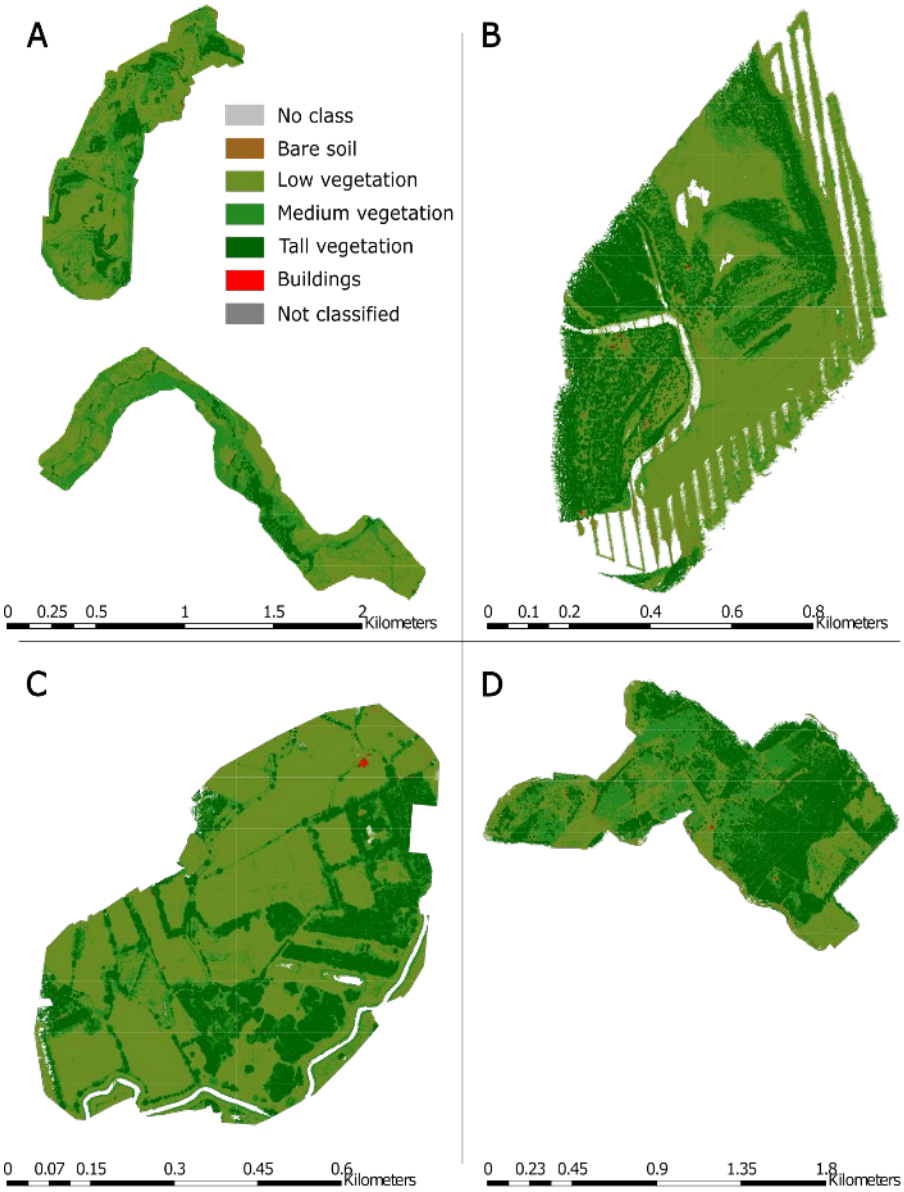
Validation

To validate the model across the different test sites, ground truth data was collected either at the same days as the drone flights or within the following week to ensure that the collected vegetation heights would be comparable with the point cloud. We estimated vegetation height at points geolocated by GNSS receivers (differential GPS devices). At each test site, we collected at least c. 100 points.

Vegetation height is complicated to measure exactly in the field. Here, we tried to estimate it at each point by recording the height where approximately half of a 20 cm imaginary horizontal line going through the recorded vegetation height was covered by the vegetation. Subsequently, these vegetation heights were translated into vegetation classes with bare soil having a vegetation height of 0 cm, low vegetation was from >0 - 25 cm, medium vegetation was from 25 - 300 cm and tall vegetation was > 300 cm. This way we could compare vegetation heights to the model output and estimate the accuracy consistently across the different test sites.

Results

A5_Fig 1 shows the classifications of our areas of interest.



A5_Fig 1 Classified point clouds (rasterized) for the test areas in (A) Understed Bakker, Denmark, (B) Örsten, Sweden, (C) Zammelsbroek, Belgium and (C) Molslaboratoriet, Denmark.

The Tables A5_Table 1,

A5_Table 2, A5_Table 3 and A5_Table 4 show the results of the validation of the deep learning vegetation classification in the four different test areas.

Appendix 5. Measuring vegetation height and cover with remote sensing tools

A5_Table 1 Validation of the deep learning vegetation classification at the test site Molslaboratoriet.

	Ground	Low Vegetation	Medium Vegetation	High Vegetation	F1-score	Recall (Correct/Total)
Ground	5	31	0	0	0.20	5 / 36 (13.9%)
Low Vegetation	8	262	12	1	0.76	262/283 (92.6%)
Medium Vegetation	0	108	274	5	0.80	274/387 (70.8%)
High Vegetation	0	2	9	14	0.62	14/25 (56.0%)
Precision (Correct/Predicted)	5/13 (38.5%)	262/403 (65.0%)	274/295 (92.9%)	14/20 (70.0%)	---	---

Overall accuracy: 555 / 731 (75.92%)

A5_Table 2 Validation of the deep learning vegetation classification at the test site Understed Bakker.

	Ground	Low Vegetation	Medium Vegetation	High Vegetation	F1-score	Recall (Correct/Total)
Ground	2	10	0	0	0.29	2 / 12 (16.7%)
Low Vegetation	0	32	4	0	0.14	32 / 36 (88.9%)
Medium Vegetation	0	29	14	2	0.43	14 / 45 (31.1%)
High Vegetation	0	0	2	3	0.60	3 / 5 (60.0%)
Precision (Correct/Predicted)	2 / 2 (100.0%)	32 / 71 (45.1%)	14 / 20 (70.0%)	3 / 5 (60.0%)	---	---

Overall accuracy: 51 / 98 (52.04%)

Table GG. Validation of the deep learning vegetation classification at the test site Örsten.

A5_Table 3 Validation of the deep learning vegetation classification at the test site Örsten.

	Ground	Low Vegetation	Medium Vegetation	High Vegetation	F1-score	Recall (Correct/Total)
Ground	0	18	0	0	0.00	0 / 18 (0.0%)
Low Vegetation	0	13	2	0	0.27	13 / 15 (86.7%)
Medium Vegetation	0	49	41	1	0.61	41 / 91 (45.1%)
High Vegetation	0	0	0	0	0.00	0/0
Precision (Correct/Predicted)	0/0	13 / 80 (16.2%)	41 / 43 (95.3%)	0 / 1 (0.0%)	---	---

Overall accuracy: 54 / 124 (43.55%)

Table BB. Validation of the deep learning vegetation classification at the test site Zammelsbroek.

A5_Table 4 Validation of the deep learning vegetation classification at the test site Zammelsbroek.

	Ground	Low Vegetation	Medium Vegetation	High Vegetation	F1-score	Recall (Correct/Total)
Ground	2	17	1	0	0.18	2 / 20 (10.0%)
Low Vegetation	0	82	1	0	0.74	82 / 83 (98.8%)
Medium Vegetation	0	39	84	3	0.76	84/126 (66.7%)
High Vegetation	0	0	10	17	0.72	17/27 (63.0%)
Precision (Correct/Predicted)	2 / 2 (100.0%)	82 / 138 (59.4%)	84 / 96 (87.5%)	17 / 20 (85.0%)	---	---

Overall accuracy: 185 / 256 (72.27%)

Discussion (1-2 pages)

Overall, this subtask showed that the deep learning based algorithm for prediction vegetation height classes in at least some cases were transferable to other landscapes in other countries. The overall accuracy of the predicted vegetation height categories is between 43 and 75 % with the highest accuracy associated with predicting medium vegetation (> 70 %).

The potential of the algorithm for transferability was shown by the fact that the overall accuracy in the original training location, Molslaboratoriet, corresponded to the accuracy found in Zammelsbroek in Belgium, almost 1000 km away. This was despite the fact that the two point clouds recorded at these two sites were recorded by both a different drone and a different sensor and moreover, also processed by different software to make the data ready for analysis (e.g., coloring the points).

In the areas where the algorithm performed more poorly, the model output was heavily influenced by the accuracy for the low vegetation class. In fact, this class was generally the hardest one to predict, yet also often the most important one for nature management as low vegetation is often an important part of the habitat quality in grasslands and wetlands (e.g. Jerrentrup *et al.* 2014). The algorithm seems to commonly confuse the low vegetation class with the medium vegetation class. More surprisingly, the algorithm also seemed to have a hard time predicting tall vegetation. This task should be easy as the differences between ground and tall vegetation is distinct and relatively large.

In our validation we collected actual vegetation heights in the field and then subsequently translated these to the vegetation height classes by binning these heights. This binning process made the field data comparable to the model output but it also likely had a large impact on the validation results. In other words, if we chose other vegetation height intervals for our bins, the validation would quite likely have yielded a different result. Thus, the validation results shown here come with significant uncertainty. However, in the case we had chosen to collect the actual vegetation height classes instead, this would also come with significant uncertainty as this is much harder to assess and collect consistently in the field. To make a fair evaluation of the algorithms abilities and usefulness in nature management one would need to also evaluate how well humans perform the task of estimating

vegetation height correctly and this was not done here. Most likely, humans will also not give the same estimates as a given reference, because vegetation height is somewhat not precisely defined, and hard to measure consistently even in the field.

All in all, we believe the methods showed potential and were surprised how well it scaled to other areas at least in some cases. However, we also believe the method needs more refining and testing before it can become operational in everyday monitoring of habitat conditions. Finally, for the validation of remotely sensed habitat condition metrics that are not exact by nature and that cannot be measured consistently in the field, we believe there is significant work ahead in developing validation methods that can give a realistic impression on how well the remote sensing methods work compared to field work.

Budget

Costs of using drone-based lidar for estimating vegetation heights depend mostly on the cost of labour, but also on the cost of equipment. To make a realistic estimate, we had to use some assumptions about these numbers in the following. These are the following.

- A drone including a suitable lidar sensor can be acquired for approximately €33,500 - €40,000.
- A drone has a lifetime of approximately 1000 flying hours, the lidar sensor itself often longer and the batteries often less.
- Software for processing the point cloud data costs approximately €2,000 per year.
- Flying a 100 ha area can be done in two-three 8-hour work days.
- Preparation will take approximately 2 hours.
- Subsequent processing and modelling of the collected data takes 1-2 work days. This could possible be automated more and hence some time could possibly be saved in an automated workflow
- Flying the drone is most effective if two persons collaborate on the task, one pilot and one that charges batteries and watches out for dangers.
- One personel costs €70 per hour
- Storage for 100 ha is roughly 50 -100 GB and this costs roughly €16 - €21

Scenario: Covering 100 ha (estimates used for calculations are the average where an interval was given above)

Drone flying :

- Drone cost per hour: €36,750 / 1000 flying hours = €36.75
- Drone cost full operation: 20 h * €36.75 = €735.00
- In total: €771.75

Drone operation:

- 2 persons for field operations: 40 h * €70 = €2,800
- 1 person for preparation: 2 * €70 = €140
- In total: €2,940

Postprocessing

- 1 person for postprocessing: 12 * €70 = €840
- Storage: €18.50
- In total: €868.50

Total costs for 100 ha: **€4,580.25**

On top of that comes the annual €2.000 per year no matter how many areas are monitored.

In a scenario where a country has e.g. 1000 nature areas of 100 ha that needs monitoring every third year, the summed costs would be approx. €1,500,000.00. This estimate is strongly conservative and the actual price of processing, logistics (combining with other drone borne habitat condition metrics and monitoring several closeby locations in the same period to save transport) and mass processing is optimized would be significantly less. The price of course also depends on how many areas need this kind of monitoring. In a future where drones are much more automated the costs would be dramatically lower. The main problem today is that the pilot needs to have visual contact with the drone at all times and this is costly in labour.

References

Hunter M O, Parente L, Ho Y-F, Bonannella C, Ferreira L G, Morton D, Consoli D, Sloat L (2025). *Global 30-m annual median vegetation height maps (2000–2022) based on ICESat-2 data and Machine Learning*. Scientific Data 12: 1470

Jerrentrup J S, Wrage-Mönnig N, Röver K-U, Isselstein J (2014). Grazing intensity affects insect diversity via sward structure and heterogeneity in a long-term experiment. Journal of Applied Ecology 51:4

Reinermann, S., Asam, S., & Kuenzer, C. (2020). *Remote sensing of grassland production and management—A review*. Remote Sensing, 12(12), 1949

Xu C, Zhao D, Zheng Z, Zhao P, Chen J, Li X, Zhao X, Zhao Y, Liu W, Wu B, Zeng Y (2023). *Correction of UAV LiDAR-derived grassland canopy height based on scan angle*. Frontiers in Plant Science 14

Appendix 6. Habitat segmentation and classification – NaturaSat approach

Partners involved:

- Mária Šibíková maria.sibikova@savba.sk (Slovak Academy of Sciences - SAS)
- Jozef Šibík jozef.sibik@savba.sk (Slovak Academy of Sciences - SAS)
- Marek Súľovský sulovsky.mark@gmail.com (Slovak Academy of Sciences - SAS)
- Martin Kelko martin2kelko@gmail.com (Slovak Academy of Sciences - SAS)
- Michal Kollár michalkollar27@gmail.com (Slovak Academy of Sciences - SAS)
- Verbesselt Sebastiaan sebastiaan.verbesselt@inbo.be (Research Institute for Nature and Forest - INBO)
- Toon Spanhove toon.spanhove@inbo.be (Research Institute for Nature and Forest - INBO)
- Rataj Jakub jakub.rataj@aopk.gov.cz (Nature Conservation Agency of the Czech Republic - AOPK ČR)
- Leština Dan dan.lestina@aopk.gov.cz (Nature Conservation Agency of the Czech Republic - AOPK ČR)
- Ježek Vít vit.jezek@aopk.gov.cz (Nature Conservation Agency of the Czech Republic - AOPK ČR)
- Gardfjell Hans hans.gardfjell@slu.se (Sveriges Lantbruksuniversitet - SLU)
- Adler Sven sven.adler@slu.se (Sveriges Lantbruksuniversitet - SLU)
- Jesper Erenskjold Moeslund jesper@ecos.au.dk

Background

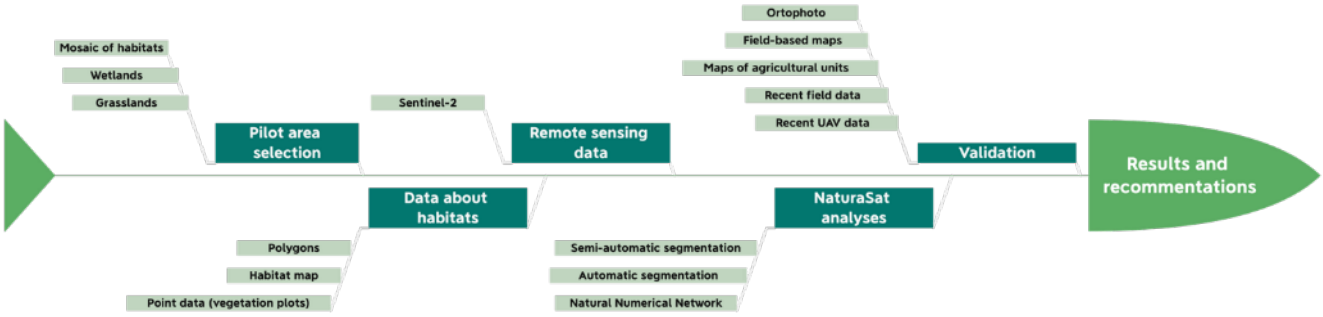
There is an accelerating need for developing improved methodological approaches to support the mapping and quality change monitoring of habitat types of conservation concern, such as Habitats Directive Annex 1 habitats, across wide areas and using remote sensing data. NaturaSat is one recently developed novel software designed for botanists, environmental researchers, and nature conservationists throughout Europe, applicable for different habitat exploration tasks (Mikula et al. 2021a). It utilizes Sentinel-2 optical data to facilitate the analysis of Natura 2000 habitats. To ensure that users across various countries can effectively utilize NaturaSat software and achieve accurate results, it is essential to conduct relevant training and testing datasets across a wide range of areas. This approach has the potential to enhance the transferability of the software tools, guaranteeing accurate identification and classification of European wetland and grassland habitats, regardless of the selected country or habitat type.

The most recent / current version of NaturaSat software aims to integrate image processing knowledge and various techniques together with vegetation science into one multipurpose tool designed to perform facilities for all requirements of habitat exploration in one place. It allows direct access to multispectral Sentinel-2 data provided by the European Space Agency. Using these data together with various vegetation databases in a user-friendly environment it supports the needs of e.g. vegetation scientists, fieldwork experts, and nature conservationists.

The NaturaSat software works with multispectral data from the Sentinel-2 satellite, UAV, and airborne imagery and with every georeferenced picture format. Implemented functionality (remote download module) allows a user to download Sentinel-2 data directly from our dedicated NaturaSat application. To access Sentinel-2 (both Level-1C and Level-2A/2Ap) data repositories, we use OpenSearch API on the Copernicus Open Access Hub.

For assessing habitat distribution, two different approaches are commonly used – object-based and pixel-based approaches, and NaturaSat software supports both. The object-based approach is represented by segmentation methods where exact habitat boundaries are found semi-automatically by expert or automatically, starting from the phytosociological relevé or other corresponding study plot whose habitat type or species assemblage is known.

Within this subtask, we piloted harmonised approaches for habitat segmentations and automatic detection of habitat presence using NaturaSat software.



A6_Fig 1 Workflow of the subtask

Recommendations

Within this subtask, we applied NaturaSat software methods for selected habitats in different partner countries – Belgium (Flanders), Czech Republic, Denmark, Sweden, and Slovakia. The task was divided into two main objectives – habitat segmentation, and habitat classification.

For segmentation, the software uses semi-automatic and automatic curve-evolving algorithms applied to Sentinel-2 data. These require an initial location estimate, typically derived from vegetation plots, existing national habitat maps or other point-based habitat observations. Users can adjust parameters related to edge detection, homogeneity thresholds, vector field filtering and stopping criteria. Segmentation results were compared with field-based habitat maps or GPS tracks using mean Hausdorff distances as quantitative metrics.

For classification, the subtask applied the Natural Numerical Network, a deep-learning method that uses statistics derived from all Sentinel-2 optical bands to classify different habitats and compute relevancy maps of their distribution. These maps indicate the likelihood that each pixel belongs to a target habitat, with higher values indicating higher relevance. Training data consisted of habitat polygons provided by partners and refined by segmentation where needed. For verification, random point samples from high- and low-relevancy areas were compared with available reference data.

The results show that NaturaSat's segmentation tools can accurately delineate habitat boundaries across all tested regions when suitable point-based occurrence data are available. Mean deviations

of one to three Sentinel-2 pixels suggest that segmentation is feasible for many habitat types, although fine-scale or heterogeneous habitats present challenges. Notable discrepancies between field-mapped and automatically segmented borders often arise in ecotones and transitional zones, where even expert delineation can be uncertain. A key advantage of automatic segmentation is its consistent inclusion or exclusion of ecotone zones compared to field mapping. The Natural Numerical Network demonstrated potential for identifying habitat presence and generating relevancy maps, given that adequate and well-distributed training data are available. This approach effectively captured habitat-specific spectral patterns, though performance was primarily influenced by the quality of the training data.

Recommendations for upscaling the methods include the determination of harmonized presets for Europe-wide habitats for segmentation methods. For upscaling the automatic detection of habitats using relevancy maps, creation of high-quality training datasets across Europe is essential.

Work description

Methods and data

Habitat segmentation

The NaturaSat software uses semi-automatic and automatic satellite image segmentation methods based on evolving planar curves (Mikula et al. 2021b,c). Both methods represent efficient and robust segmentation tools when an “initial estimate” of the desired area is available.

This availability varies among EU countries. In the case of Natura 2000 habitats, the point-wise estimate of habitat occurrence is often available in vegetation databases (Šibík 2012, Chytrý et al. 2016) or national habitat maps with different precision. Thus, the software tools allow focusing the Sentinel-2 image to the selected habitat occurrence data point, e. g. phytosociological relevé, or to the user-defined area, and then allow the user to perform semi-automatic or automatic segmentation by evolving the initial curve, either in the form of a straight line (semi-automatic segmentation) or in the form of a small circle (automatic segmentation).

The final mathematical model is given by the corresponding nonlinear intrinsic partial differential equation which is discretized and solved numerically by the flowing finite volume method (Mikula & Ševčovič 2001, Mikula et al. 2010).

Since variability of habitats in Europe is enormous, NaturaSat offers great variety of settings, including:

- Edge detector (Allows setting the sensitivity parameter for detecting the edges in the data)
- Vector field filtration type (Allows choosing between three types of filtrations for vector field: Linear, Nonlinear, and Geometric-MCF).
- Homogeneity detector for automatic segmentation with more options of calculating “homogenous habitat area” including EPS function (Allows setting the percentage share of homogeneous pixels around the mean value in the selected area. For example, the pixel will be considered homogeneous if it falls within an interval of 20% around the mean value of the selected area.
- Segmentation settings as variation in time-step, total number of steps or auto-stop

- Complex list of settings can be found in user guide (http://www.algoritmysk.eu/wp-content/uploads/2025/02/NaturaSat_user_manual.pdf)

Habitat classification

The pixel-based approach is represented by relevancy maps based on a new deep learning method called Natural Numerical Network (the method is presented in Mikula et al. 2021). The vector of computed values in all optical bands (mean, maximal, minimal, standard deviation) creates pixel characteristics that are used by the natural numerical network to classify the pixel within a habitat with certain relevancy (for details Mikula et al. 2021).

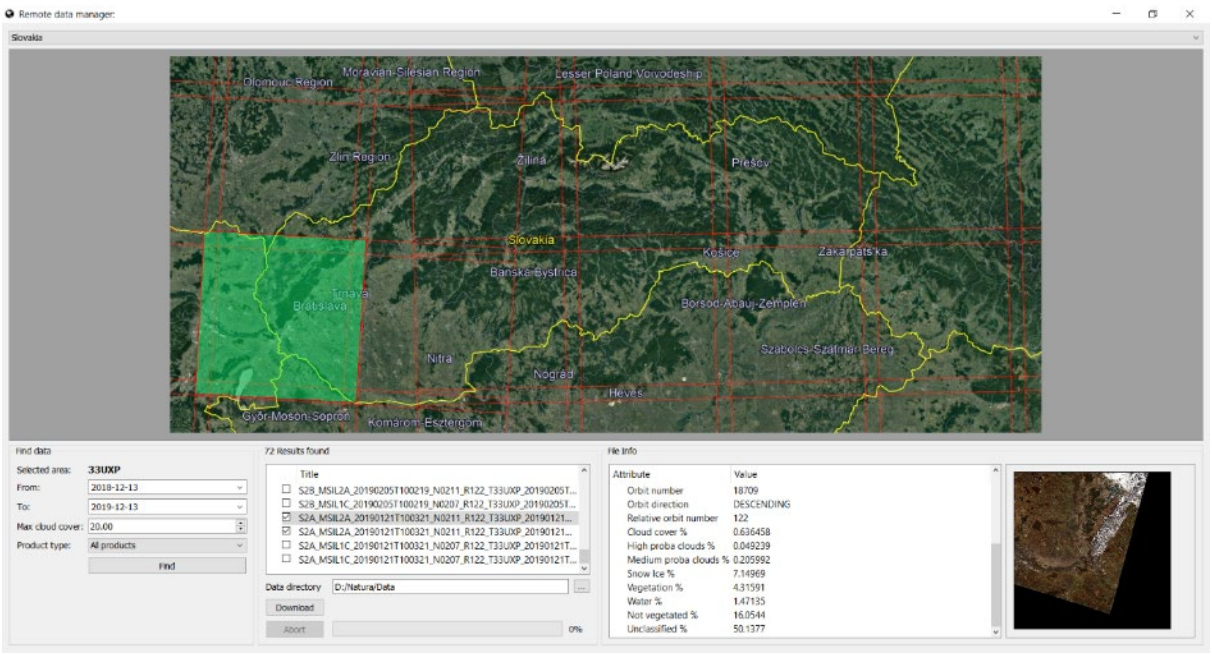
Natural numerical networks represent deep learning algorithm based on the numerical solution of nonlinear forward-backward diffusion equations on complete graphs. Forward diffusion causes the movement of points in a feature space toward each other. The opposite effect, keeping the points away from each other, is caused by backward diffusion which yields the desired classification. The natural numerical network contains just a few parameters that are optimized in the learning phase of the method. After learning parameters and optimizing the topology of the network graph, classification necessary for habitat identification is performed.

The method produces a so-called relevancy map in which the greyscale intensity of every pixel represents habitat occurrence relevancy. This approach allows detecting habitats in degraded conditions (medium or low relevancy should show habitats with fewer diagnostic species). It is also sensitive to transitional habitat types that have high relevancy values for two different habitats.

Data acquisition

Segmentation and classification of habitats within the Biodiversa+ project was mainly tested on Sentinel-2 data due to their free availability for the whole Europe (although NaturaSat software also supports other formats such as ortophotomaps or drone images). The remote download module allows a user to download Sentinel-2 data directly from the application. To access Sentinel-2 (both Level-1C and Level-2A/2Ap) data repositories, we use OpenSearch API on Copernicus Open Access Hub.

Appendix 6. Habitat segmentation and classification – NaturaSat approach



A6_Fig 2 User interface of the NaturaSat software allowing direct download of Sentinel-2 data

Polygon data representing ground truth was provided by partners from Czech Republic, Belgium and Slovakia. In Czech republic data from field mapping done by experts in botany from the years 2016-2023 was provided. In Slovakia, field mapping was used to obtain the gps tracks of selected habitats. In Belgium (Flanders), the map of water bodies was used as a ground truth data.

For habitat classification, training data were obtained using following steps:

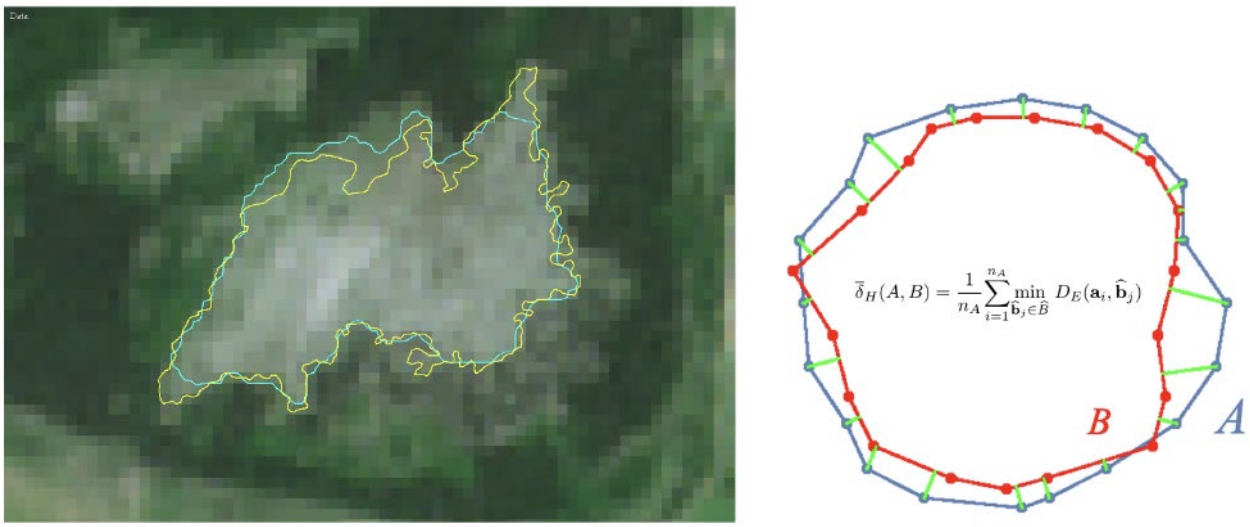
- Local habitat map was provided by partners
- Different provided formats (mostly .shp) were converted into .kml file and divided into individual curves
- Habitats smaller than 9 pixels, and thinner than 2 pixels were excluded due to the Sentinel-2 resolution limitation
- Geographic stratification was done to cover equally the variability of vegetation in selected region
- Sentinel-2 data from selected region was downloaded, period close to the date when ground data were collected
- Selected curves (habitat polygons) often represent parcels or administrative units, so the habitat does not occur in the whole polygon area. Thus, the borders were tuned to actual shape by using automatic segmentation in NaturaSat (evolve function)
- Resulting polygons were used for the training phase of the Natural Numerical network

Verification

(A) Segmentation verification

After performing the automatic segmentation of habitats in Sentinel-2 data, we compared the segmentation result with the GPS tracks or field-based maps obtained by field experts. For this quantitative comparison of two curves, we used the classical (maximal) Hausdorff distance (https://en.wikipedia.org/wiki/Hausdorff_distance) and the so-called mean Hausdorff distance, see

e.g. (Krivá et al., 2010), which are general tools for computing distance of curves, surfaces and even more complicated geometrical continuous or discrete objects (sets).



A6_Fig 3 Example of gps track (yellow) and automatic segmentation (blue) and the formula of Hausdorff distance used for comparison of two curves

(B) Classification verification

After computation of relevancy maps for selected habitats, 100 random points were created in high relevancy areas (white areas with probability of habitat occurrence more than 90%), and 100 random points from low relevancy areas. In Slovakia these points were compared with habitat map obtained in the field during vegetation season 2024 and 2025. In Denmark, random points were compared with the map of fields that includes also information about intensity of grassland use.

Results

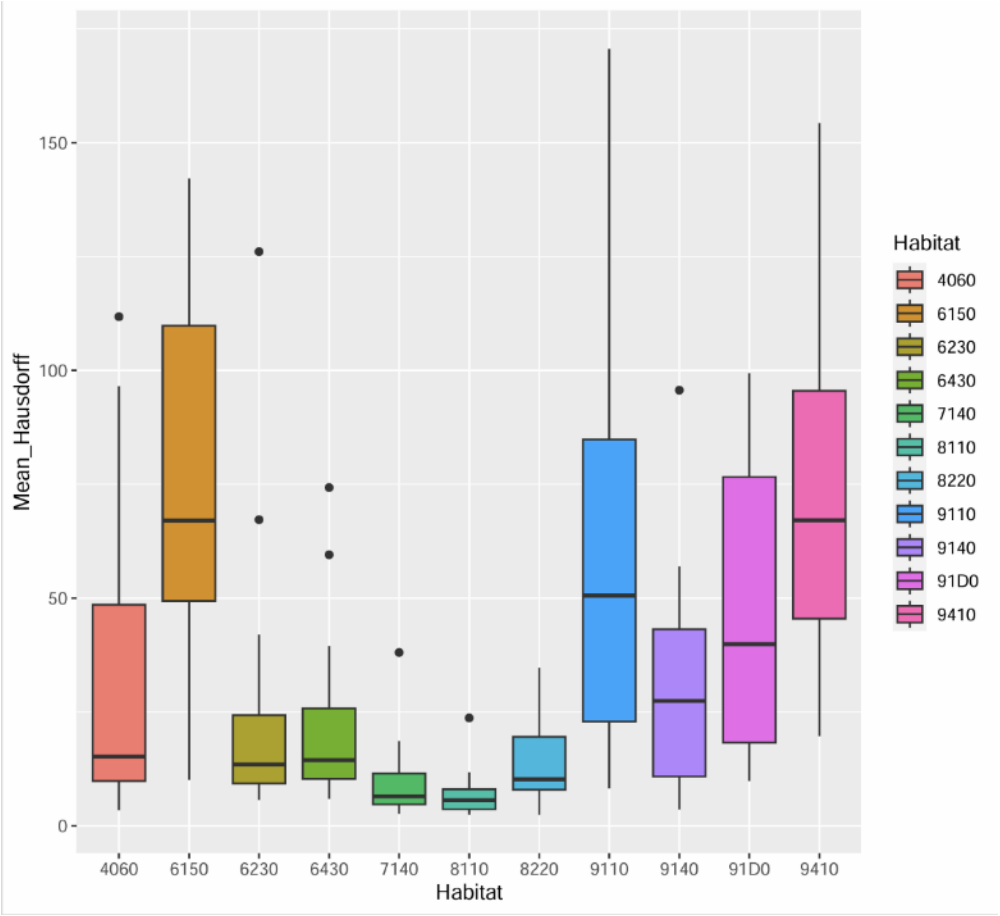
Habitat segmentation

Czech Republic

In Czech republic, 11 Annex 1 habitat types were automatically segmented in NaturaSat software:

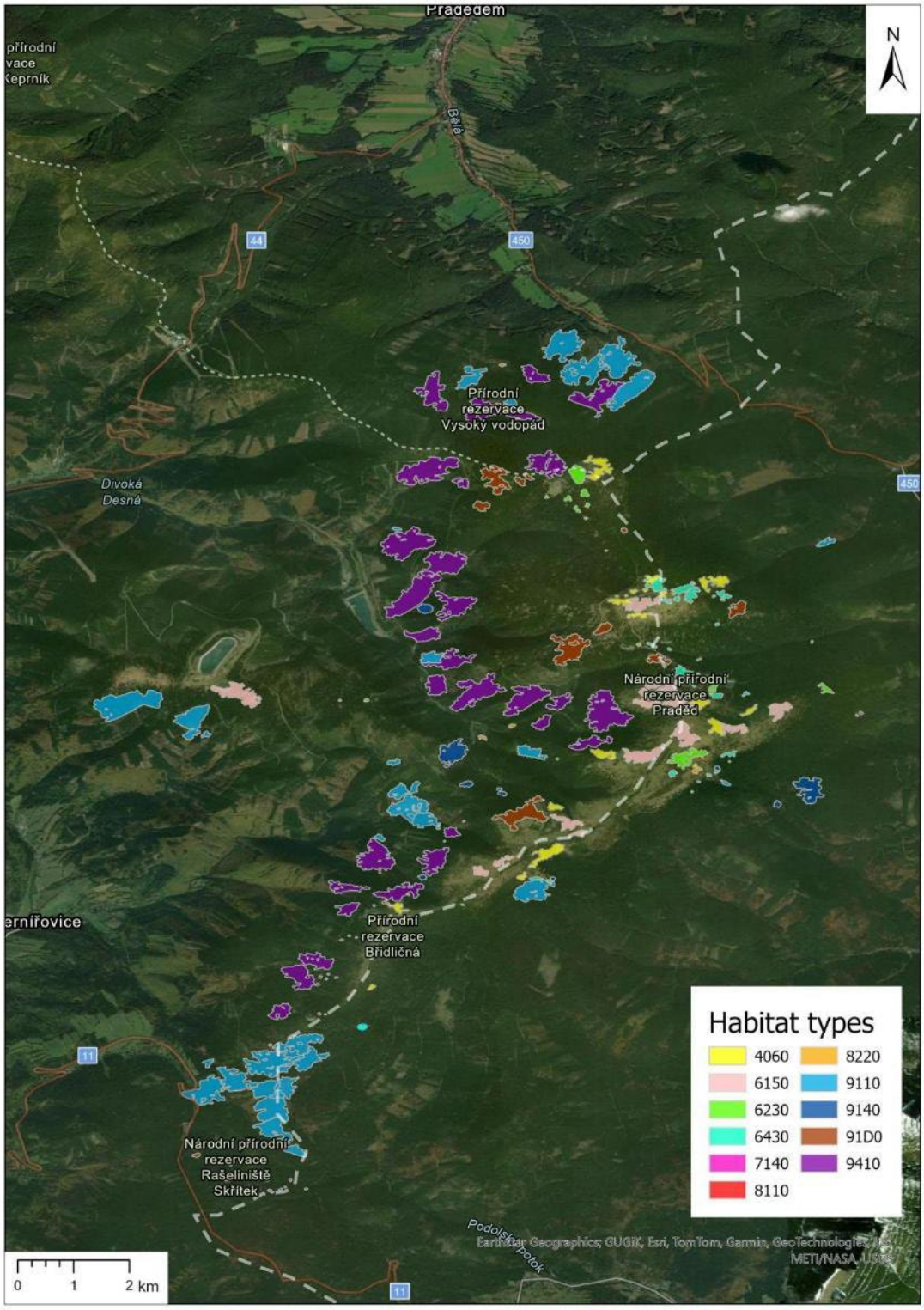
- 4060 Alpine and Boreal heaths
- 6150 Siliceous alpine and boreal grasslands
- 6230 Species-rich Nardus grasslands, on silicious substrates in mountain areas (and submountain areas in Continental Europe)
- 6430 Hydrophilous tall herb fringe communities of plains and of the montane to alpine levels
- 7140 Transition mires and quaking bogs
- 8110 Siliceous scree of the montane to snow levels (*Androsacetalia alpinae* and *Galeopsietalia ladani*)
- 8220 Siliceous rocky slopes with chasmophytic vegetation
- 9110 Luzulo-Fagetum beech forests
- 9140 Medio-European subalpine beech woods with *Acer* and *Rumex arifolius*

- 91D0 : Bog woodland
- 9410 : Acidophilous Picea forests of the montane to alpine levels (*Vaccinio-Piceetea*)



A6_Fig 4 Mean Hausdorff distance between field habitat maps and automatically segmented habitats.

The most accurate results were obtained in case of small-scale, rare habitats like 7140 Transition mires and quaking bogs, 8110 Siliceous scree of the montane to snow levels and 8220 Siliceous rocky slopes with chasmophytic vegetation. The differences between field data and segmentation were less than 10 meters, which means less than a pixel resolution of Sentinel-2 data. The lowest accuracy was found in forests and 6510 meadow habitat, which could be also caused by large habitat extent.

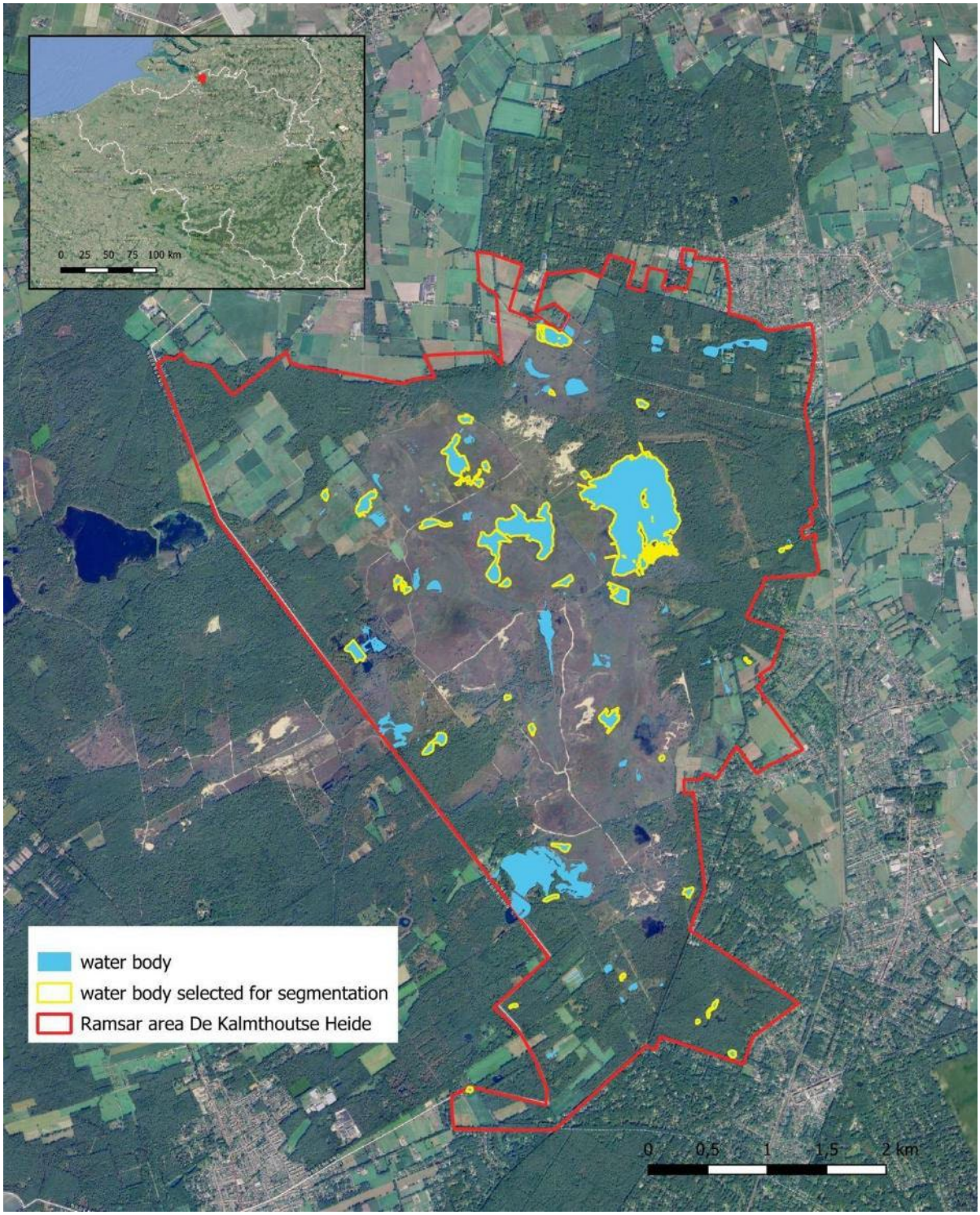


A6_Fig 5 The area of Praded nature reserve and distribution of segmented habitats.

Belgium

For Belgium, water habitats were chosen, because exact borders exist from field mapping. (the maps are available at https://zenodo.org/records/14203168). 35 water bodies in Ramsar site Kalmthoutse

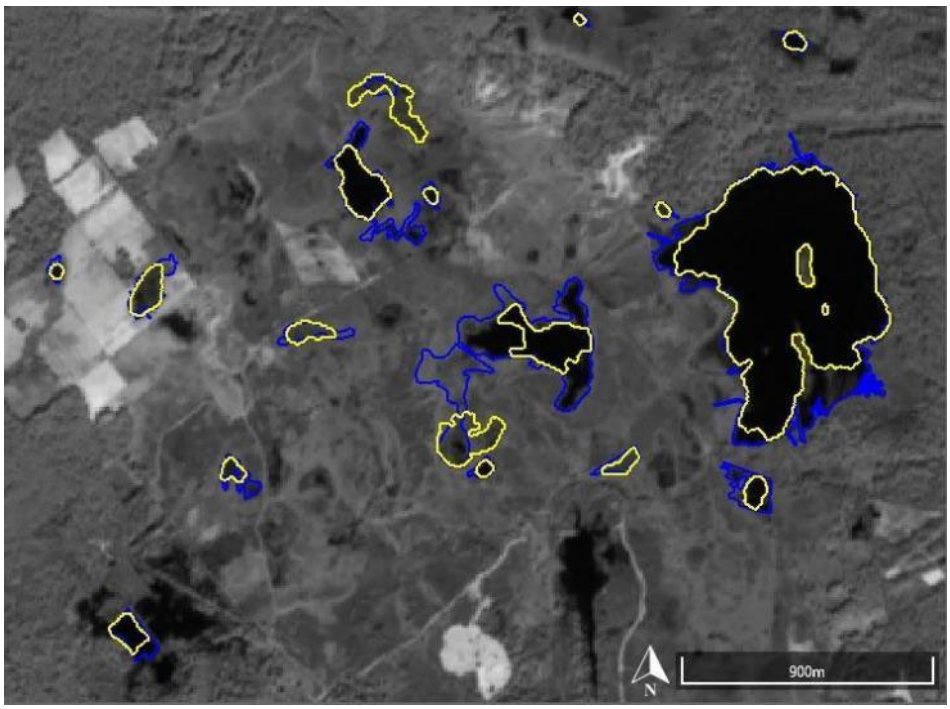
Heide were selected and automatically segmented using NaturaSat. Segmentations were done at Sentinel-2 dataset from 2025-03-20, NIR or Red band was used for segmentation. EPS was set to 20% and two time-steps of automatic segmentation were computed.



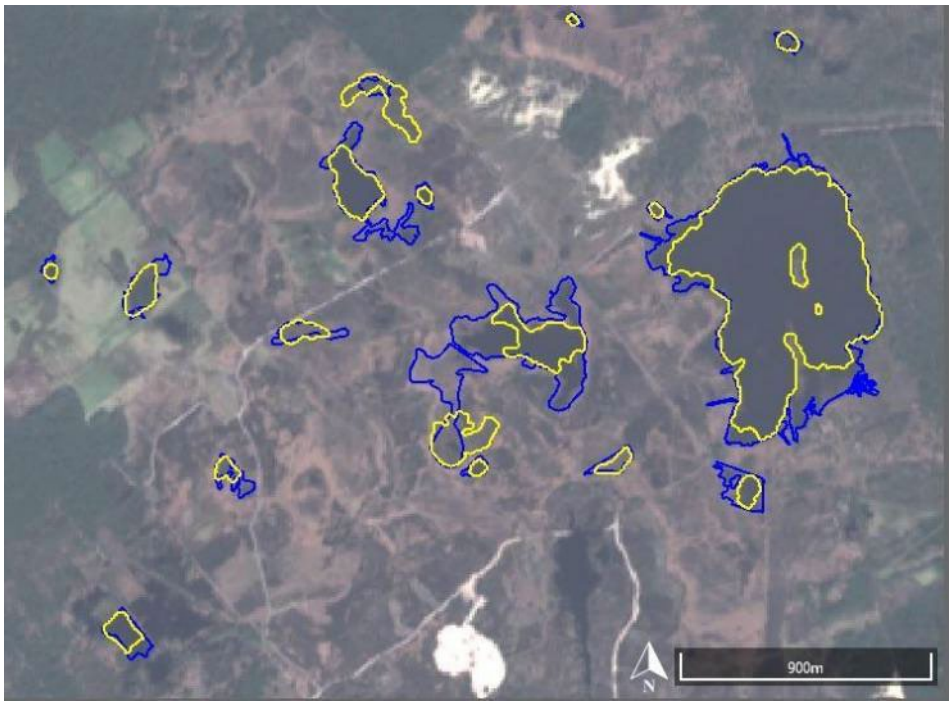
A6_Fig 6 The area of Kalmthoutse Heide Ramsar site and occurrence of segmented habitats.

Mean Hausdorff distance was 29m (3 pixels), and the differences between original borders and segmented one was mainly caused by different structures at the borders of segmented areas, e.g.

reed beds. During field mapping, reed beds were classified as part of water bodies, whereas automatic segmentation typically includes only water habitats.



A6_Fig 7 Original map of water bodies (blue) and automatic segmentation results (yellow) visualized in NIR band of Sentinel-2 data



A6_Fig 8 Original map of water bodies (blue) and automatic segmentation results (yellow) visualized in RGB band combination of Sentinel-2 data

Slovakia

In Slovakia, 2 grassland habitat types were automatically segmented in NaturaSat software:

- 5130 *Juniperus* communis formations on heaths or calcareous grasslands
- 6190 Rupicolous pannonic grasslands (*Stipo-Festucetalia pallentis*)

5130 represents sparsely to densely developed *Juniperus communis* thickets on mostly calcareous substrate. In many places, these are successive overgrown thermophilic pastures, formed by vegetation of the *Festuco-Brometea* class. The species composition of the herb layer is mainly influenced by the involvement density of the species *Juniperus communis*.

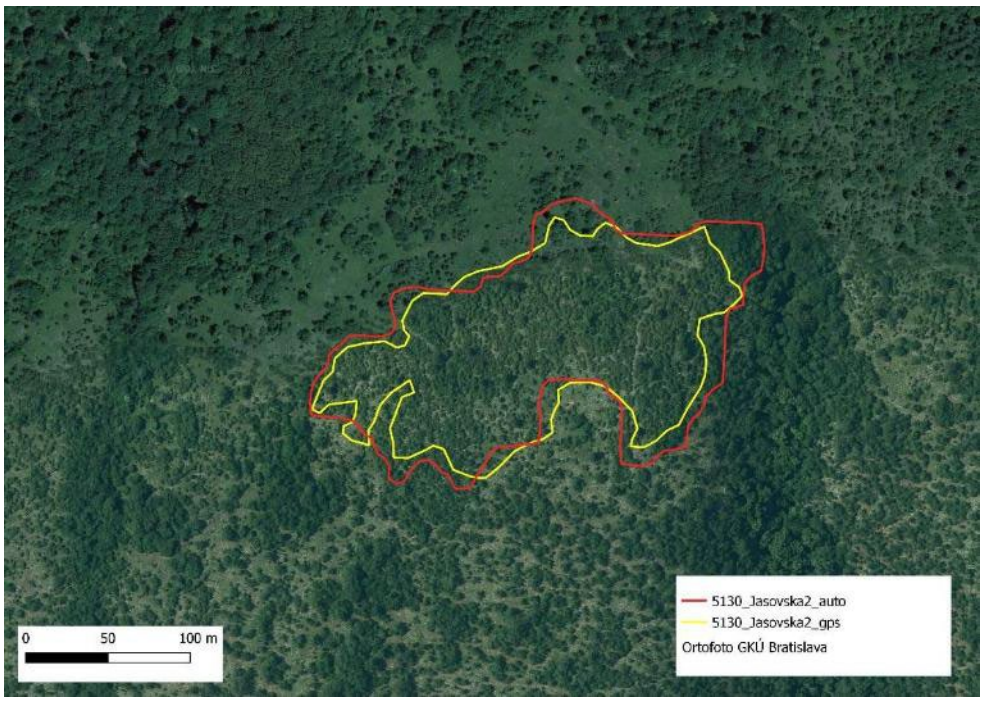
A6_Table 1 Mean Hausdorff distance between automatic segmentation and gps tracks of the habitat 5130

Curve name	Automatic versus GPS
5130_DlhaVes1_auto	12.0298
5130_DlhaVes1_gps	
5130_DlhaVes2_auto	15.3213
5130_DlhaVes2_gps	
5130_Jasovska1_auto	9.94318
5130_Jasovska1_gps	
5130_Jasovska2_auto	8.0744
5130_Jasovska2_gps	
5130_Paklan_auto	6.03781
5130_Paklan_gps	
5130_SilickaJablonica1_auto	13.9625
5130_SilickaJablonica1_gps	
5130_SilickaJablonica2_auto	5.97096
5130_SilickaJablonica2_gps	
Average	10.1914

For segmentation of this habitat, Sentinel 2 images were the most suitable during the non-vegetation period, when contact species of trees are not yet leafed and stands of evergreen juniper are easier to distinguish. On the other hand, a problem arises in the transitional contact areas, where the juniper is in the undergrowth of the dominant higher trees. As they are not yet leafy, segmentation can also include areas that can already be marked as forest. In the case of transitional contact areas, a reduction of the parameter EPS (% from mean) to 3% appears to be effective.

The minimum size of the square as the initial curve in the automatic segmentation for the default setting is 40x40m. This presents a problem of generating an initial curve with sufficient

homogeneity in the case of irregular small-area or linear occurrences. Creating the initial circular curve often encounters the same problem as a square. In addition, creating a circular curve is relatively difficult for smaller areas, as it is drawn on an unexpected area after the initial click and drag. The solution would be to draw the initial curve as a rectangle, respectively as an irregularly shaped polygon.



A6_Fig 9 GPS track (yellow) and automatic segmentation (red) of the habitat 6190 in Slovenský karst protected area.

Habitat type 6190 involves dry grasslands on limestone and are usually exposed to solar irradiation, on steep (often south-facing) slopes with very shallow soils. The vegetation cover is mostly unclosed or patchy.

A6_Table 2 Mean Hausdorff distance between automatic segmentation and gps tracks of the habitat 6190

Curve name	Automatic versus GPS
6190_Dlhyvrch_202106_auto	5.03308
6190_Dlhyvrch_202109_auto	6.04399
6190_Dlhyvrch_gps	
6190_Gombasek_202106_auto	5.74202
6190_Gombasek_202109_auto	5.65613
6190_Gombasek_gps	
6190_Hornyvrch1_202106_auto	6.31847
6190_Hornyvrch1_202109_auto	6.17245

6190_Hornyvrch1_gps	
6190_Hornyvrch2_202106_auto	5.85675
6190_Hornyvrch2_202109_auto	6.37686
6190_Hornyvrch2_gps	
6190_Hornyvrch3_202106_auto	5.28359
6190_Hornyvrch3_202109_auto	5.87037
6190_Hornyvrch3_gps	
6190_Okruhle_202106_auto	8.55098
6190_Okruhle_202109_auto	6.81599
6190_Okruhle_gps	
6190_Plesivske1_202106_auto	8.44611
6190_Plesivske1_202109_auto	7.68704
6190_Plesivske1_gps	
6190_Plesivske2_202106_auto	16.4993
6190_Plesivske2_202109_auto	18.3541
6190_Plesivske2_202110_auto	8.94412
6190_Plesivske2_gps	
6190_Slavecke_202106_auto	18.7364
6190_Slavecke_202109_auto	14.2095
6190_Slavecke_202110_auto	10.8351
6190_Slavecke_gps	
6190_Turna_202106_auto	7.05596
6190_Turna_202109_auto	7.03894
6190_Turna_202110_auto	6.59913
6190_Turna_gps	
6190_Zadielska_202106_auto	4.67946
6190_Zadielska_202109_auto	5.37899
6190_Zadielska_gps	
6190_Zemne_202106_auto	13.5887
6190_Zemne_202109_auto	6.68735
6190_Zemne_gps	
Average	8.4615

This habitat occurs primarily in dry and rocky sites, which are dried out after the summer at the end of the vegetation season compared to adjacent more mesophilic grassland communities. For this reason, it appears in most cases as a more accurate basis for automatic segmentation of the Sentinel-2 image from September (in some cases, October).

This type of vegetation is also characterized by a variety of herb cover depending on the amount of exposed bedrock. So in some cases, better segmentation results were obtained when the parameter EPS (% from mean) was increased to 10.

The minimum size of the square as the initial curve in the automatic segmentation for the default setting is 40x40m. This involves the problem of generating an initial curve with sufficient homogeneity in the case of irregular small-area or linear occurrences. As discussed earlier, also here c reating the initial circular curve often encounters the same problem as a square. In addition, creating a circular curve is relatively difficult for smaller areas. The solution would be to draw the initial curve as a rectangle, respectively as an irregularly shaped polygon.

In general, we can conclude that habitat boundary detection reached pixel accuracy in the case of every tested habitat (A6_Table 1). Problem areas were in the case of forests the so-called “ecotone zones” – the areas of dense shrubs along the forest boundary, which is often composed of the same species as the forest itself, but younger individuals. Field experts sometimes classify these zones subjectively. The advantage of using NaturaSat software is demonstrated by the fact that software is consistent in including or excluding “ecotone zones”. Another common difference against GPS tracks is the inclusion of small paths or thin streams into habitat areas. This phenomenon is caused by the Sentinel-2 data resolution (10x10 meters), where objects narrower than 10 meters are hard to detect. Such elements are often considered a natural part of the habitat in forest habitats, so their inclusion is acceptable.

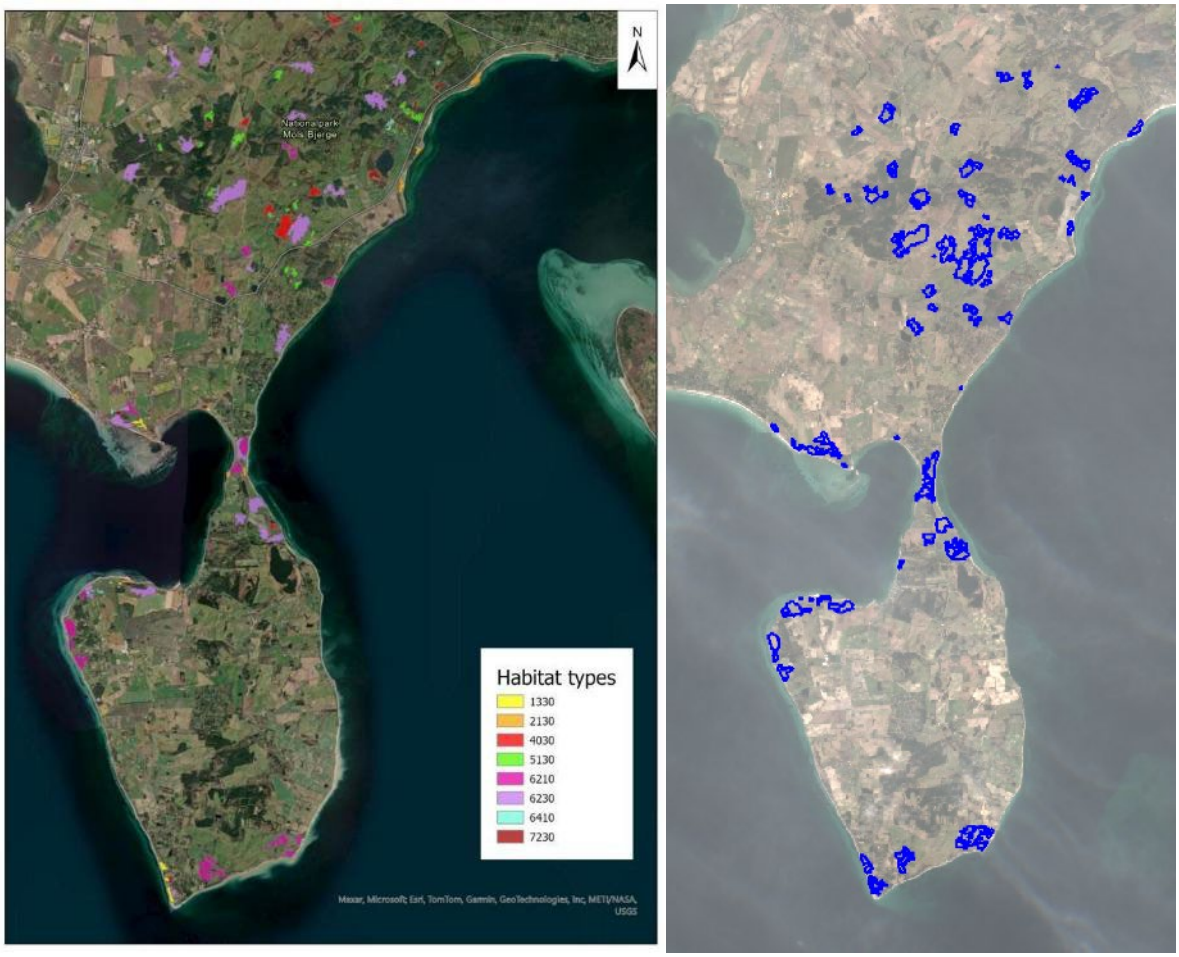
The seasonal variability of habitats seems to be a crucial parameter that must be considered when the Sentinel-2 dataset is chosen for segmentation, especially in case of grassland habitats.

Habitat classification

Denmark

In Denmark, following Annex 1 habitats were provided by project partners and segmented for the training phase:

- 1330 Atlantic salt meadows (Glauco-Puccinellietalia maritimae)
- 2130 Fixed coastal dunes with herbaceous vegetation ('grey dunes')
- 4030 European dry heaths
- 5130 *Juniperus communis* formations on heaths or calcareous grasslands
- 6210 Semi-natural dry grasslands and scrubland facies on calcareous substrates (*Festuco-Brometalia*)
- 6230 Species-rich Nardus grasslands, on silicious substrates in mountain areas (and submountain areas in Continental Europe)
- 6410 *Molinia* meadows on calcareous, peaty or clayey-silt-laden soils (*Molinion caeruleae*)
- 7230 Alkaline fens



A6_Fig 10 Map of habitats in Denmark (left) and visualization of segmented training polygons (right).

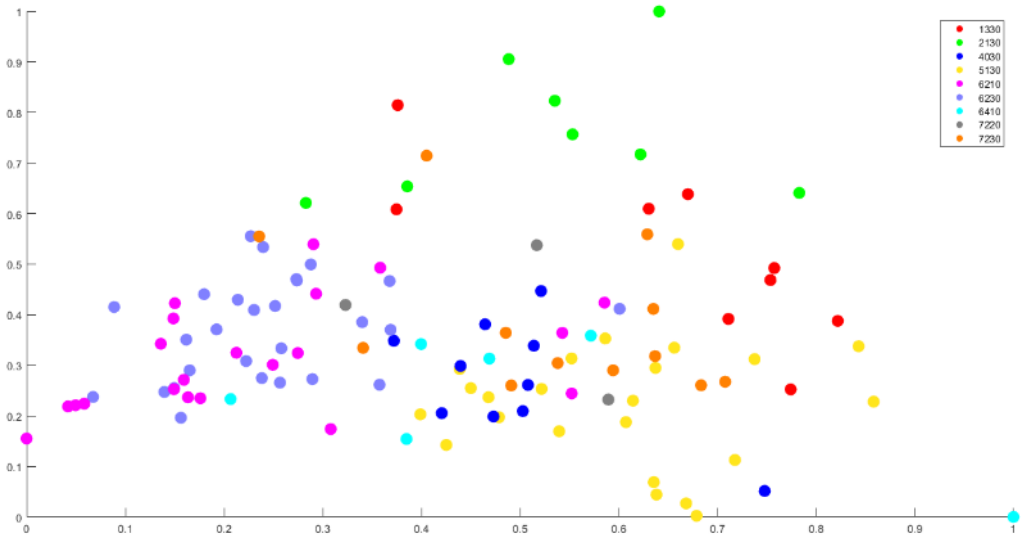
After the inspection of the dataset, some habitats that were too small, or represented by a low number of locations were excluded.

A6_Table 3 shows the habitat types in the examined area and the number of polygons in each habitat.

A6_Table 3 Habitat codes, number of polygons and number of squares

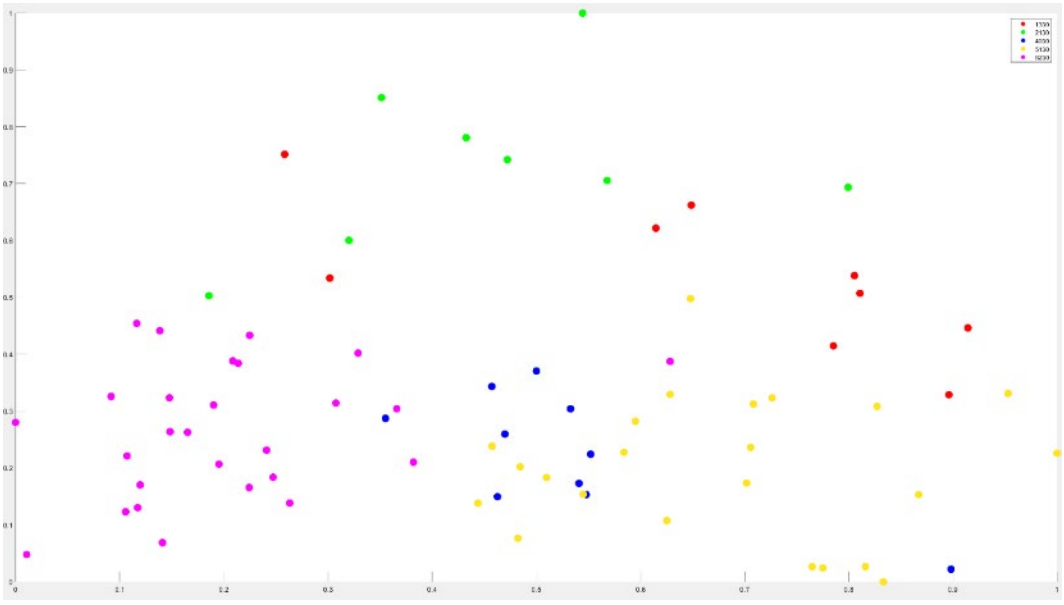
Habitats codes	Number of polygons	Number of squares LDS119x60	Number of squares LDS77x60	Number of squares LDS35x60
1330	10	9	9	7
2130	10	8	8	7
4030	10	10	10	7
5130	30	23	23	7
6210	27	21	-	-
6230	30	27	27	7
6410	10	6	-	-
7220	5	3	-	-
7230	12	12	-	-
Sum:	144	119	77	35

The NatNet model requires constructing representative squares in the polygons. Such methodology is proposed to ensure the regularity of the features across the training samples. In our case, 7x7-pixel squares are used, and they are randomly placed in the interior of the polygons. Table 1, in the third column, shows the number of squares in each habitat, and one can see that the number of squares differs from the number of polygons. It is caused by the fact that some polygons are smaller than the 7x7 square. The feature space used in the learning process is constructed from statistical characteristics of the non-derived Sentinel-2 bands and the derived NDVI index. The first learning dataset, LDS119x60, consisted of 119 representative squares from 9 habitats, each in 60-dimensional feature space. To reduce the dimension of the feature space and maximise data variance of the data, PCA was used.



A6_Fig 11 Visualised LDS119x60 in 2 dimensions, reduced by PCA.

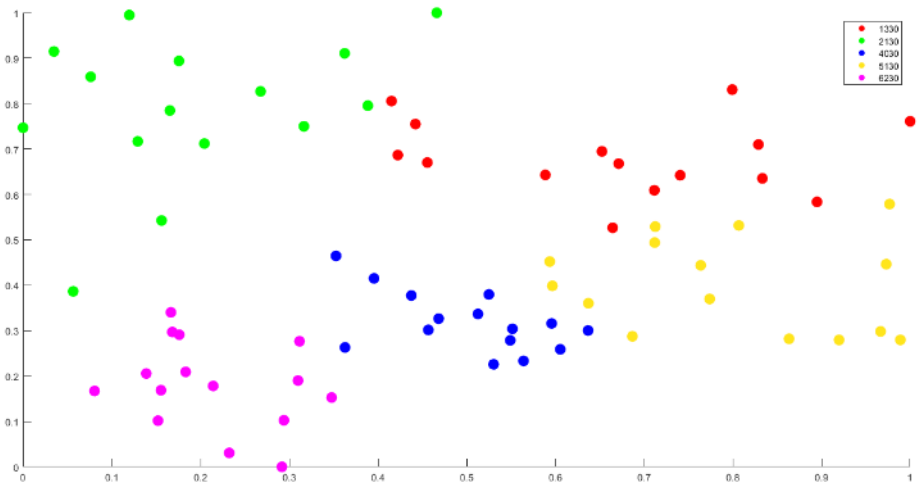
A6_Fig 11 shows the LDS119x60 in 2 dimensions, where the spectral characteristics of samples were mixed. As a result, the learning process wasn't released. Based on the LDS119x60 visualisation, the number of habitats enrolled in the classification process was reduced. Some habitats (e.g. habitat 6410 and 7220) were excluded because of the low number of representative squares and because their positions after PCA transformation were inconsistent. These changes imply the construction of the new learning dataset LDS77x60. Details of the dataset composition are shown in 3, and the visualisation of LDS77x60 in 2 dimensions after PCA reduction is shown in Figure 12.



A6_Fig 12 Visualised LDS77x60 in 2 dimensions, reduced by PCA.

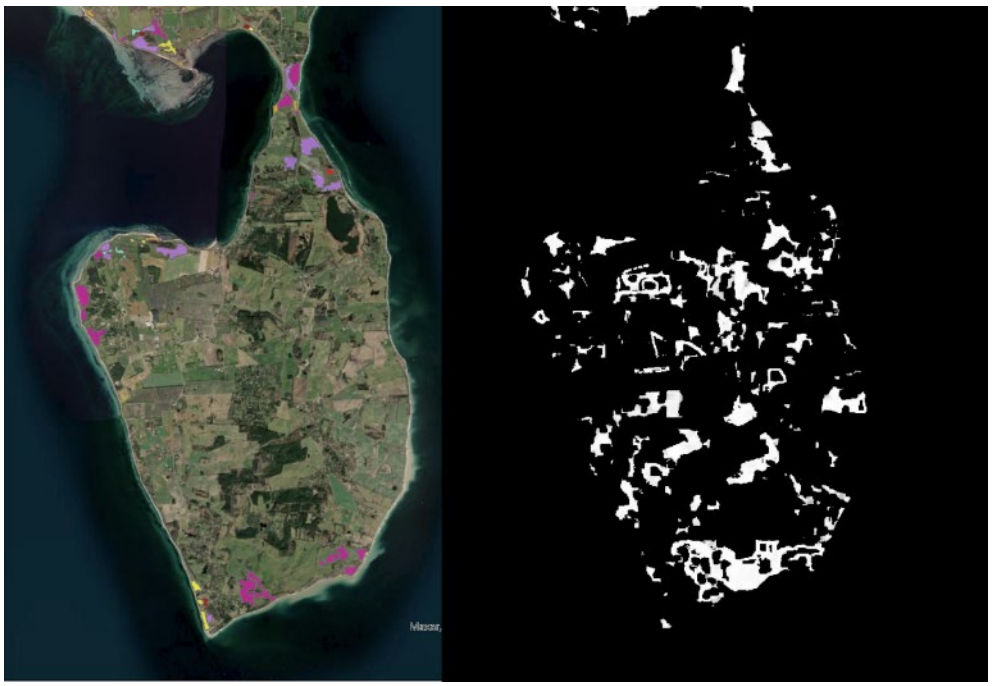
To ensure stratified sampling across habitats, the new learning dataset LDS35x60 is constructed by selecting only seven representative squares from each habitat type. Table 3, in the fifth column,

shows the composition of the LDS35x60, and *Figure 13* visualises the dataset with 2 dimensions of feature space after PCA reduction.

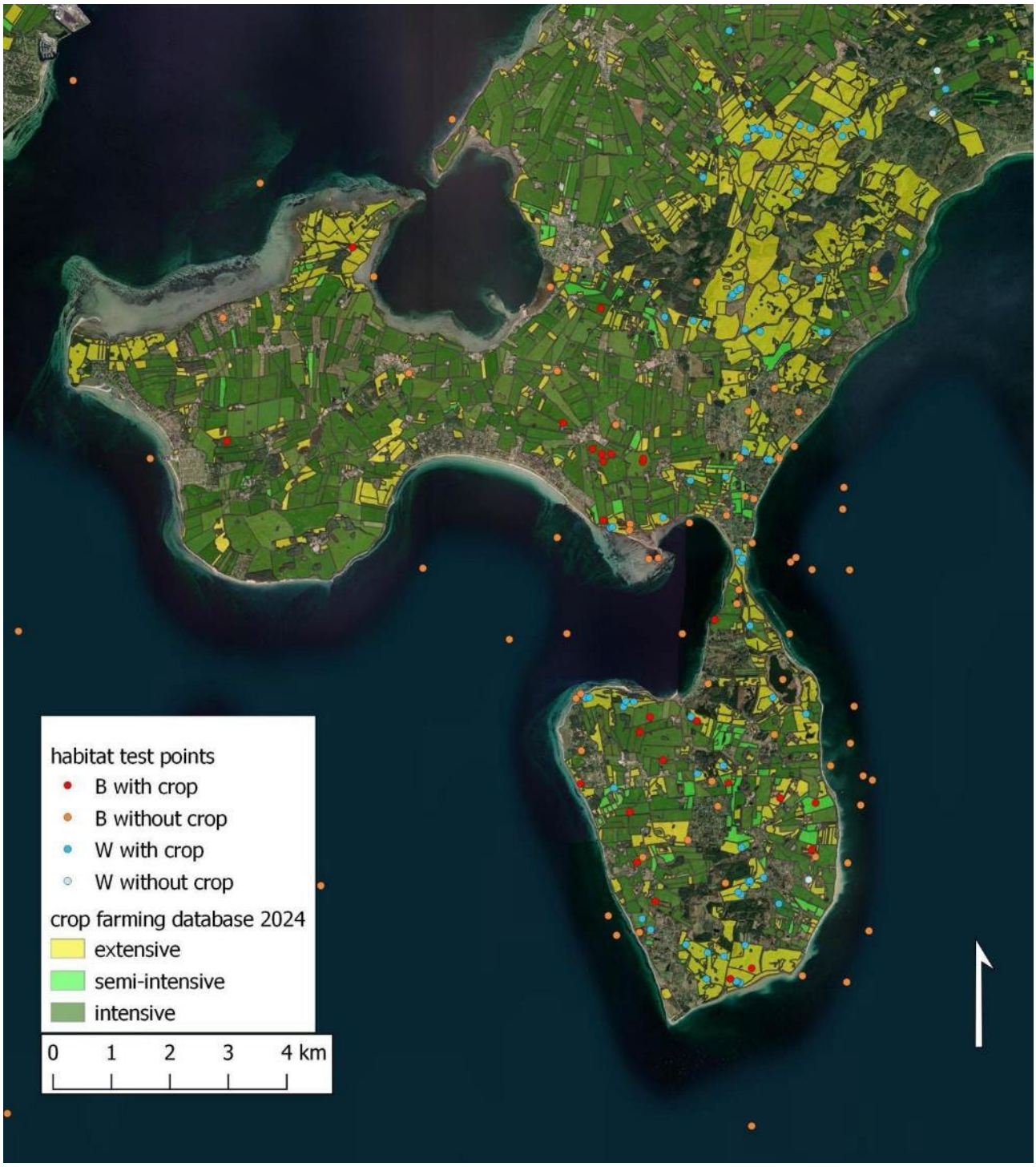


A6_Fig 13 Visualised LDS35x60 in 2 dimensions, reduced by PCA.

The learning process was launched for LDS35x60, aiming to find model parameters that yield the highest possible accuracy. The training accuracy for LDS35x60 is 94.29%. After successful learning phase a relevancy map was created (Fig. 14). One hundred random points were created in high relevancy areas (white areas with probability of habitat occurrence more than 90%), and a similar amount of random points in low relevancy areas (Fig. 15).



A6_Fig 14 . Relevancy map for the habitat complex 6230/6210



A6_Fig 15 Random points and the crop farming database that was used for validation

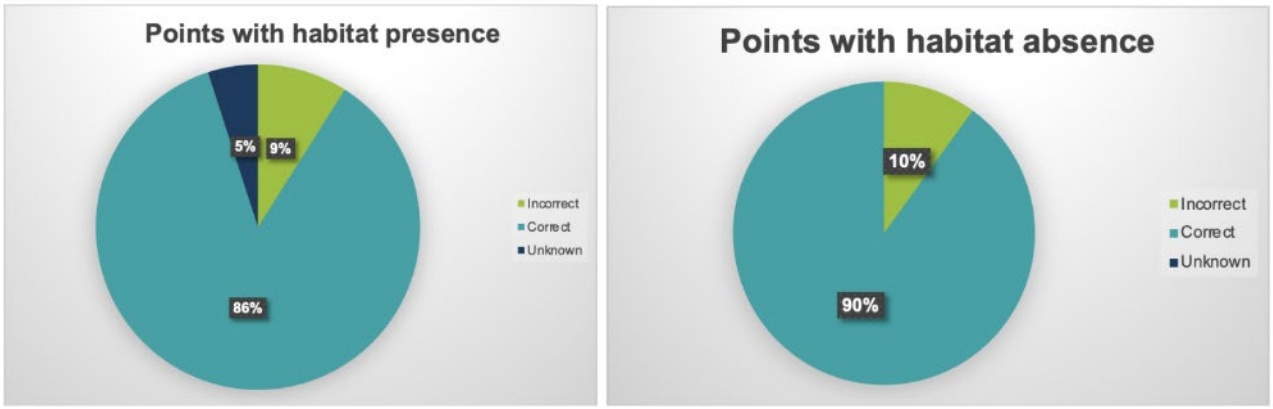
A6_Table 4 Categories from the crop farming database where the random points with high habitat probability occurs

Spring barley	1
Winter barley	1
Winter wheat	2
Winter rye	2
Winter hybrid rye	3
Permanent grass, very low yield	2
Permanent grass, low yield	1
Permanent grass, standard yield	1
Environmental grass AECM scheme 2 (0 N) (permanent)	23
Permanent grass, under 50% clover/alfalfa	1
Grass with clover/alfalfa, under 50% legumes (rotation)	1
Grass and clovergrass without N norm, under 50% clover (rotation)	1
Grass under 50% clover/alfalfa, low yield (rotation)	1
Recreative purpose	2
Permanent grass and clovergrass without N norm, under 50% clover	9
Fallow with mowing	3
AECM no selection, not farmland	40
Set aside for polinators	1

A6_Table 5 . Categories from the crop farming database where the random points with low habitat probability occurs

AECM no selection, not farmland	2
Environmental grass AECM scheme 2 (0 N) (permanent)	2
Fallow with mowing	2
Grass and clovergrass without N norm, under 50% clover (rotation)	1
Green grain of winter wheat	1
Peas	2
Permanent grass, standard yield	1
Permanent grass, very low yield	1
Set aside for polinators	1
Spring barley	3

Winter hybrid rye	2
Winter rape	2
Winter wheat	6
no-grassland	74



A6_Fig 16 Percentage of correctly classified points

Sweden

The learning process for the NatNet was based on polygon data collected from south eastern Sweden, using data from the TUVA database (https://jordbruksverket.se/e-tjanster-databaser-och-appar/e-tjanster-och-databaser-stod/tuva) . The studied sites are located in the provinces Blekinge and Öland. They contain a large number of different habitat types, and the codes used for the habitats correspond to the Annex I habitat types used in the assessment of the European Habitats directive. A6_Table 6 illustrates the different habitat types in the examined area along with the corresponding number of polygons for each habitat.

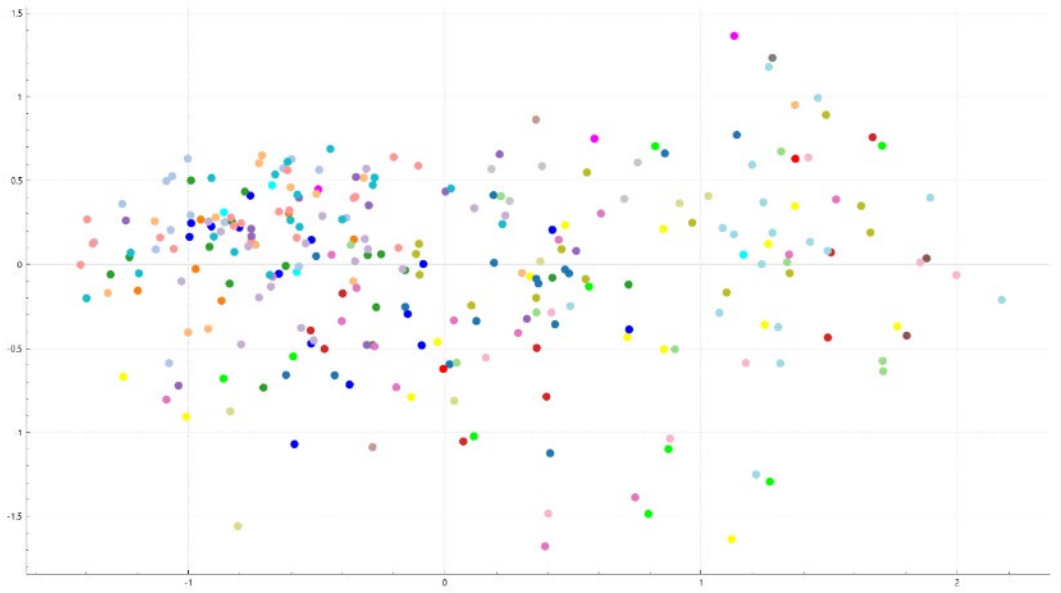
A6_Table 6 Habitat codes, number of polygons and number of squares in each learning dataset.

Habitats codes	Number of polygons	Number of squares LDS268x60	Number of squares LDS77x60	Number of squares LDS63x60	Number of squares LDS60x60
1310	5	2	-	-	-
1630	14	9	-	-	-
2130	20	15	15	15	14
2330	4	4	-	-	-
4010	3	3	-	-	-
4030	21	14	-	-	-
5130	24	16	16	16	16
6110	19	15	15	15	15
6120	7	6	-	-	-
6210	18	14	-	-	-
6211	18	16	-	-	-
6230	15	9	-	-	-
6270	20	9	-	-	-
6280	21	18	-	-	-
6410	20	12	-	-	-
6411	24	18	-	-	-
6430	8	2	-	-	-
6450	2	2	-	-	-
6510	25	15	-	-	-
6530	11	8	-	-	-
7140	1	1	-	-	-
7210	7	5	-	-	-
7230	20	14	14	-	-
8230	19	6	-	-	-
8240	19	18	-	-	-
9070	22	17	17	17	15

Sum:	387	268	77	63	60

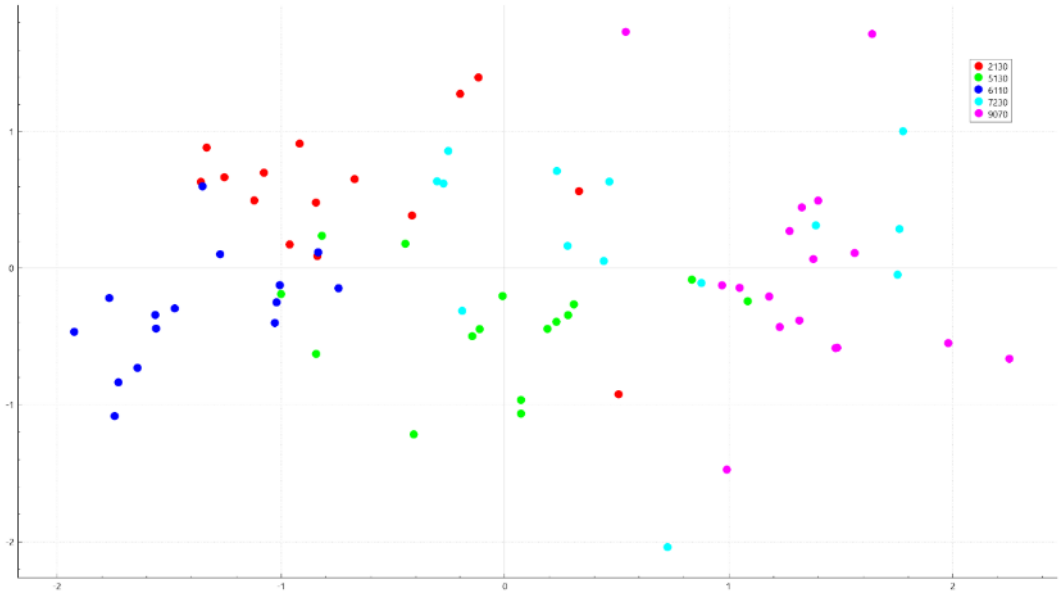
The NatNet model requires constructing representative square patches in polygons to ensure the regularity of features across training samples. In our workflow, we aim to extract 11x11-pixel squares because they provide sufficient spatial context for NatNet to learn stable local patterns. However, when a polygon is too small to fit this size, the square dimensions are adaptively reduced to 9x9 or 7x7 pixels. The squares are then randomly positioned in each polygon, which helps capture diverse local variations of habitat.

Table 6, in the third column, shows the number of squares in each habitat, and we can see that the number of squares differs from the number of polygons. It is caused by the fact that some polygons are even smaller than a 7x7 square. The feature space used in the learning process is constructed calculating statistical characteristics of the 14 Sentinel-2 bands and the derived NDVI index. The research began with a learning dataset, LDS268x60, consisting of 268 representative squares from all habitats in the Tuva region, each in a 60-dimensional feature space. To reduce the dimension of the feature space and maximise data variance of the data, PCA is used.



A6_Fig 17 Visualised LDS268x60 in 2 dimensions (2 principal components), reduced by PCA

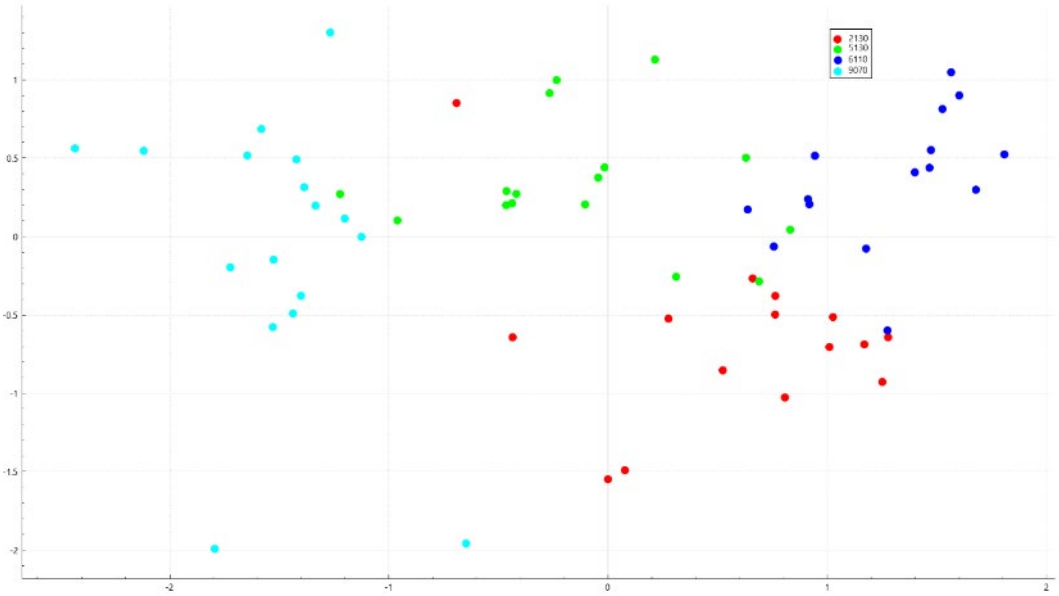
A6_Fig 17 shows the LDS268x60 in 2 dimensions, where these 2 dimensions correspond to the 2 principal components after PCA reduction. Analysis of the figure shows that the samples are thoroughly mixed. Following this analysis, the learning process was unreasonable to release for all habitats together, so the dataset was split and five habitats were chosen, namely 2130, 5130, 6110, 7230, and 9070. As a result, the new LDS77x60 is constructed (A6_Fig 18), and the number of representative squares is shown in Table 6.



A6_Fig 18 Visualised LDS77x60 in 2 dimensions (2 principal components), reduced by PCA.

The visualisation of the LDS77x60, in A6_Fig 18, shows that the 7230 habitat correlates with the features of the 9070, 2130, and 5130 habitats. Because the goal is to use a predefined feature space, the 7230 habitat was excluded from classification. However, this provides a space for future research on adjusting the feature space composition.

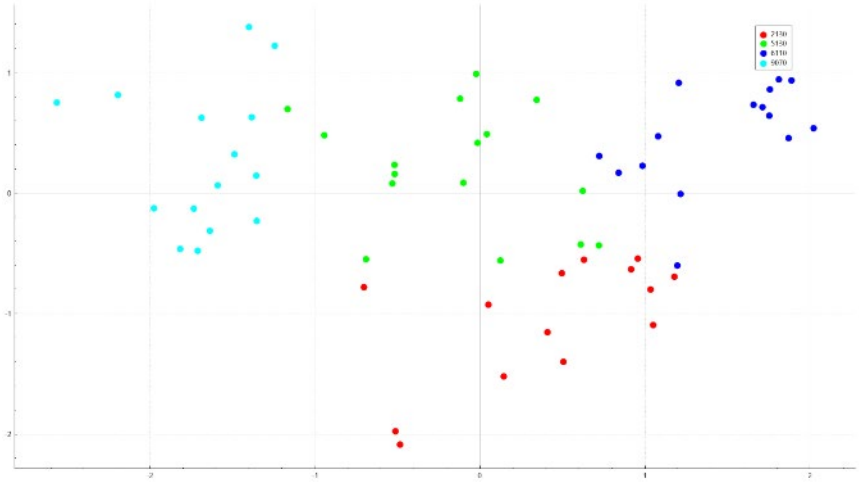
At the end, we have a learning dataset LDS63x60 with four habitats, and A6_Table 6in the fifth column shows the constellation of the dataset. The habitat visualisation is shown in A6_Fig 19.



A6_Fig 19 Visualised LDS63x60 in 2 dimensions (2 principal components), reduced by PCA

The LDS63x60 dataset seems appropriate for learning the NatNet model. Upon reviewing the visualisation in Figure 3, it becomes apparent that two samples stand out as potential outliers, as

they are positioned far from the 9070 cluster after PCA reduction. One sample is located between the 5130 and 9070 clusters, which indicates it could be misclassified. To prevent this and ensure more accurate learning of the NatNet model, the final LDS60x60 dataset will exclude these three representative squares from the polygons.



A6_Fig 20 Visualised LDS60x60 in 2 dimensions (2 principal components), reduced by PCA.

The learning dataset LDS60x60 is visualised in A6_Fig 20, and its composition is shown in the last column of Table 1. The learning process for LDS60x60 is launched to find model parameters that yield the highest possible accuracy. The training accuracy for LDS60x60 was 95%.

A6_Fig 21 The visualization of 6110 habitat (one component in the Alvar complex habitat type) on Öland island.

Slovakia

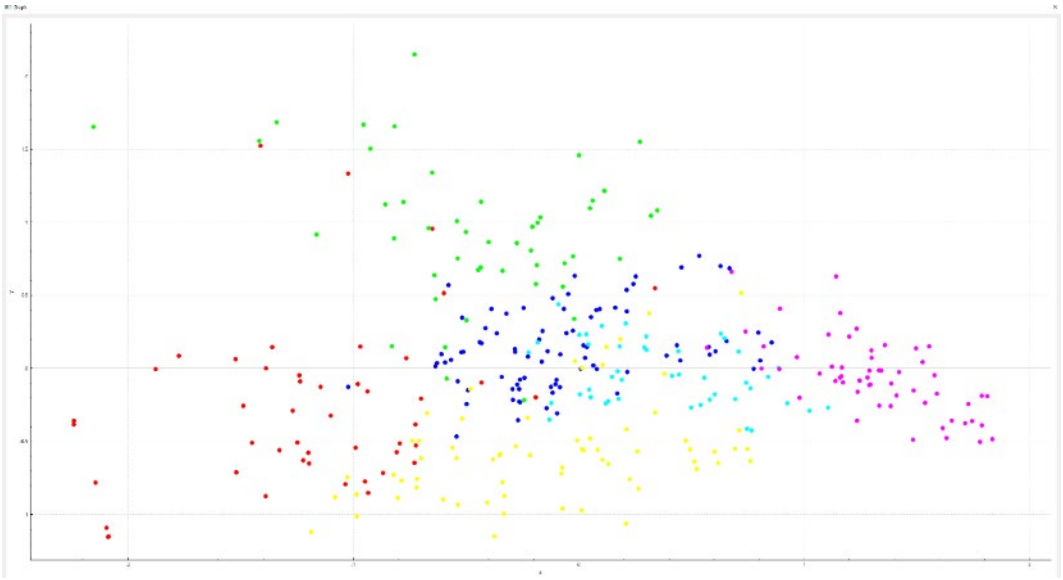
In Slovakia, we focused on the wet meadow habitat 6440, specifically the alluvial meadows of river valleys of the *Cnidion dubii*. However, since these meadows often create a mosaic with adjacent forest habitats, we also included these forests in the training phase. A6_Table 7 presents the habitat types included in this study, along with the number of polygons for each habitat.

A6_Table 7 Habitat codes, number of polygons and number of squares.

Habitats codes	Number of polygons	Number of squares LDS344x52	Number of squares LDS287x52	Number of squares LDS240x52
91E0 Alnus	18	47	41	40
91E0 Salix	22	47	41	40
91F0	33	83	65	40
91G0	31	44	41	40
9110	31	57	48	40
6440	16	66	51	40
Sum:	151	344	287	240

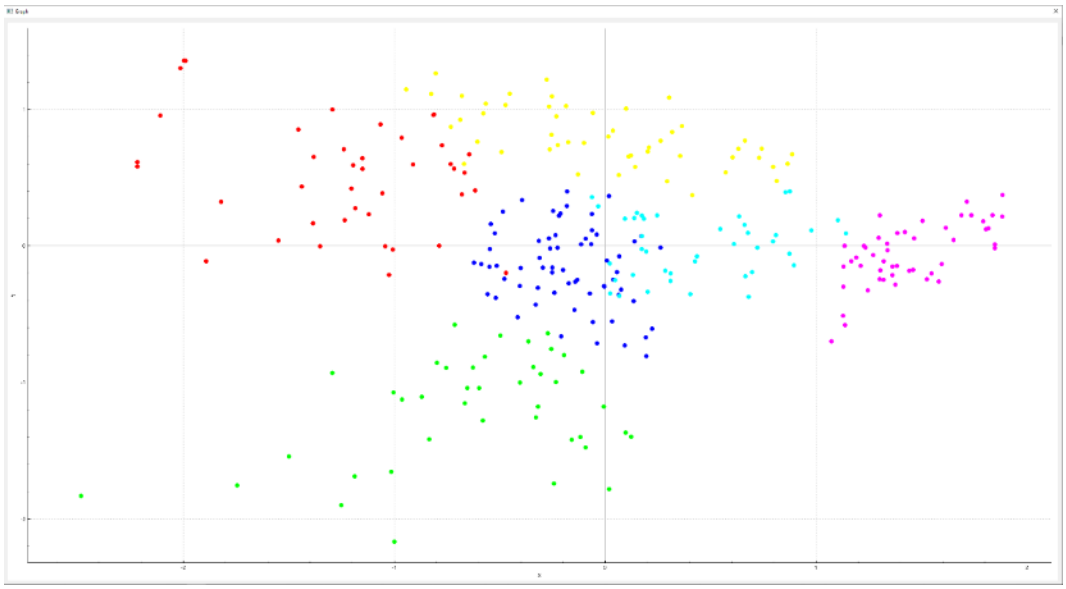
To preserve uniformity in feature extraction, the NatNet model requires that square patches be constructed for each polygon. In our workflow, we construct 11x11-pixel squares because they offer sufficient spatial context for NatNet to learn stable local patterns. However, to avoid excluding any polygons at the beginning of the classification process, we adaptively reduce the square size for smaller polygons, using 9x9 or even 7x7 pixels when necessary. This ensures that every polygon, regardless of its dimensions, can still contribute training information. In this study, we explore an alternative workflow for constructing the learning dataset.

Typically, the dataset is created by placing one or two squares inside each polygon, and after reviewing the training samples, these squares are either adjusted or removed. In contrast, our initial learning dataset, LDS344x52, includes as many squares as can reasonably fit in each polygon. After reviewing, we will reduce the number of squares in the dataset. Table 7, in the third column, shows the number of squares in each habitat for LDS344x52, and one can see that the number of squares in each habitat exceeds the number of polygons. The feature space used in the learning process is constructed from statistical characteristics of the non-derived Sentinel-2 bands and the derived NDVI index. In this study, the feature space will have 52 dimensions, and PCA will be used to reduce the dimension of the feature space for visualisation and maximise the data variance of the data.



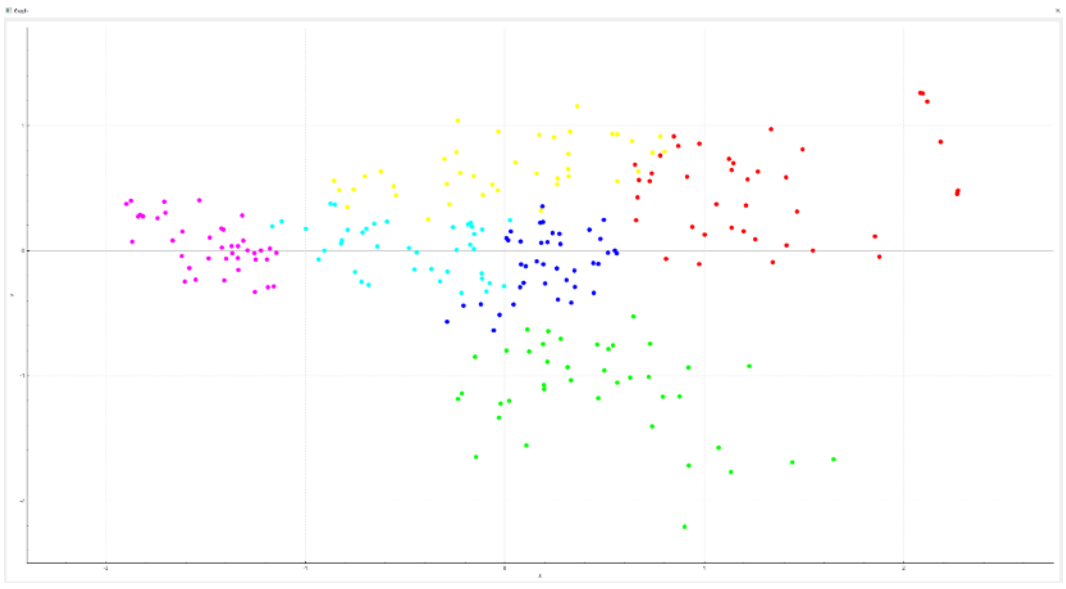
A6_Fig 22 Visualised LDS344x52 of target habitats (see Table 7) in 2 dimensions, reduced by PCA.

A6_Fig 22 shows the LDS344x52 in 2 dimensions, and we can notice that samples from the same habitat form clusters, but there are also overlapping samples. According to our strategy, mixed samples were excluded from the learning dataset, and the LDS287x52 was constructed. Details of the dataset composition are shown in *Table 7*, in the fourth column, and the visualisation of LDS287x52 in 2 dimensions after PCA reduction is shown in A6_Fig 23. One can see that clusters are clearer, and the learning process can achieve a high success rate.



A6_Fig 23 Visualised LDS287x52 of target habitats (see Table 7) in 2 dimensions, reduced by PCA.

However, the learning dataset, LDS287x52, does not have a uniform distribution of samples across habitat types. To establish a uniform sampling, a new dataset, LDS240x52, was constructed by selecting 40 representative squares from each habitat type. This stratified construction ensures balanced habitat coverage in the learning set. A6_Fig 24 visualises the LDS240x52 dataset with 2 dimensions of feature space after PCA reduction. For the LDS240x52 dataset, the learning process was conducted to optimise the model parameters that maximise classification performance. The model attained a training accuracy of 92.1%. This success rate demonstrates that the LDS240×52 dataset provides an adequate foundation for effective classification of the new samples.



A6_Fig 24 Visualised LDS240x52 of target habitats (see Table 7) in 2 dimensions, reduced by PCA.

Discussion

Natura 2000 sites are a network of protected areas across Europe, established to conserve the most valuable and threatened habitats and species. Natura 2000 habitats refer to the specific habitats within this network, which are ecologically significant and typically consist of various vegetation units characterized by different structures and defined by their species composition, reflecting environmental conditions. This complexity can make accurate recognition challenging. In natural landscapes, these habitats often form a mosaic, frequently interspersed with ecotone zones or extensive transitional areas between different habitats. Consequently, even experienced experts may have differing opinions on habitat boundaries. Therefore, adopting a harmonized approach is crucial for effective habitat mapping on larger scales, especially when non-experts are involved.

Segmentation methods implemented in NaturaSat software (Mikula et al., 2021a) are based on evolving curve approach that was successfully used in different fields of image-processing, including medical data, and the accuracy of border detection was widely tested (e.g. within ESA PECS 5, 7 and 8 projects). These methods give us the opportunity to identify habitat borders consistently, and in comparison, with field tracking also independently on terrain complications.

Segmentation methods in NaturaSat software are ready-to-use in all European countries where the basic requirement, existence of point data of habitat occurrences (e.g. vegetation plots) is fulfilled. The software offers a variety of settings for specific cases such as habitats that include heterogenous structures (Juniper shrubs). Default settings in NaturaSat software meet the basic requirements for the majority of habitats with relative homogenous structure. The segmentation can be fitted to the selected habitat by choosing the best edge detector function, homogeneity function, expanding force and many other variables.

To establish harmonized habitat mapping across Europe, future tuning of available settings will be very useful. Moreover, creation of some “presets” with the best recommended parameters settings for different habitats shall be included in the software. Requirement for the initial point information about habitat presence could be easily obtained from the European Vegetation Archive (EVA) database where vegetation plots from whole Europe are collected, thus the segmentation methods could be successfully adopted in other countries.

Use of Sentinel-2 data has some objective limitations, mainly the 10x10m pixel resolution, which is too coarse for some rare, small-scale habitats. However, these data are freely available, could be downloaded directly through the NaturaSat software and offer a dense time-series. Thus, final choice of the data can follow the phenology of target habitats.

Within this subtask, we were able to reach the accuracy 1-3 pixels of the mean hausdorff distance between field data and automatic segmentation in all tested countries (Czech Republic, Belgium and Slovakia). Some larger values of the maximal Hausdorff distance are caused by the “ecotone zones”. Such transition zones are usually hard to classify, even subjectively by expert knowledge. Nevertheless, we see strong potential for the identification of habitat borders by using the segmentation methods and Sentinel-2 data. It allows a necessary simplification that helps determine the sharp edges in the images and avoid a subjective bias.

Conclusions and recommendations

To improve habitat segmentation within the Natura 2000 framework, it is essential to verify segmentation results through quantitative comparisons with field data, such as GPS tracks and expert field maps. Utilizing the classical Hausdorff distance metrics for evaluating the accuracy of segmentation allows for a clear assessment of how closely automated results align with physical boundaries.

In the Czech Republic, accurate segmentation was achieved for small-scale and rare habitats, such as transition mires and quaking bogs, where the differences between field data and segmentation were minimal. Conversely, the accuracy diminished for larger habitats like forests and meadows, highlighting the need for improved methods in these areas.

In Belgium, the successful segmentation of water habitats demonstrated the importance of precise field mapping to establish exact borders, while differences in structural classifications at boundaries emphasized the challenges faced during automatic segmentation. In Slovakia, the segmentation of grassland habitats revealed that the timing of satellite imagery acquisition is critical. Using Sentinel-2 images during non-vegetation periods yielded better results for identifying Juniperus communis formations, while careful adjustment of parameters like EPS helped refine results in transitional areas. The challenges of working with small or irregularly shaped habitats necessitate the development of more flexible initial curve drawing methods, such as allowing for rectangular or irregular polygons to enhance segmentation processes.

Overall, achieving pixel-level accuracy in habitat boundary detection is feasible for various tested habitats, although challenges persist in ecotone zones and areas with dense shrub layers. The consistent application of NaturaSat software aids in the objective inclusion or exclusion of these zones, reducing subjective bias. Given the limitations of Sentinel-2 data resolution, it is crucial to account for seasonal variability and select appropriate imaging dates for effective segmentation, particularly for grassland habitats. In conclusion, ongoing refinement of segmentation methods, validation against field data, and consideration of habitat seasonal dynamics are vital for improving habitat classification accuracy.

To enhance habitat classification, it is essential to develop a harmonized classification system that minimizes discrepancies among experts in recognizing habitat boundaries and defining ecotone zones. Continuing to utilize and refine the segmentation methods in NaturaSat software, including the creation of “presets” for different habitat types based on empirical data, will facilitate user-friendly application. Promoting the use of the vegetation data from various database, e.g. European Vegetation Archive (EVA) (Chytrý et al. 2016) to gather initial point data on habitat occurrences will ensure comprehensive datasets for effective habitat mapping. Additionally, exploring higher-resolution satellite imagery can improve classification accuracy, particularly for small-scale habitats. Very accurate training datasets that represent homogeneous, even smaller polygons of target habitats are crucial for creating relevancy maps. Analyzing satellite images from various dates throughout the year is necessary when creating features that distinguish similar habitat types from each other. Future learning datasets should be constructed with uniform sampling across habitat types to enhance classification performance and address imbalances. Fostering collaboration among countries to share best practices and findings, including joint workshops, is also vital. Regularly updating and optimizing model parameters based on the latest data will help maintain high classification accuracy while validating against field data.

In conclusion, a standardized approach to habitat classification is crucial for improving accuracy and collaboration among experts within the Natura 2000 framework. The segmentation methods in NaturaSat software demonstrate significant potential for effectively identifying habitat borders, while the accessibility of Sentinel-2 data provides valuable resources for habitat mapping. Constructing robust learning datasets is essential for maximizing model performance. Continued research into classification methodologies and ecological relationships among habitats will further enhance conservation strategies across Europe.

References

Chytrý, M., Hennekens, S. M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., Landucci, F., Schaminée, J. H. J., Aćić, S., Agrillo, E., Ambarlı, D., Angelini, P., Apostolova, I., Attorre, F., Berg, C., Bergmeier, E., Biurrun, I., Botta-Dukát, Z., Brisse, H., Campos, J. A., Carlón, L., Čarni, A., Casella, L., Csiky, J., Ćušterevska, R., Dajić Stevanović, Z., Danihelka, J., De Bie, E., de Ruffray, P., De Sanctis, M., Dickoré, W. B., Dimopoulos, P., Dubyna, D., Dziuba, T., Ejrnæs, R., Ermakov, N., Ewald, J., Fanelli, G., Fernández-González, F., FitzPatrick, Ú., Font, X., García-Mijangos, I., Gavilán, R. G., Golub, V., Guarino, R., Haveman, R., Indreica, A., Işık Gürsoy, D., Jandt, U., Janssen, J. A. M., Jiroušek, M., Kącki, Z., Kavgacı, A., Kleikamp, M., Kolomiychuk, V., Krstivojević Ćuk, M., Krstonošić, D., Kuzemko, A., Lenoir, J., Lysenko, T., Marcenò, C., Martynenko, V., Michalcová, D., Moeslund, J. E., Onyshchenko, V., Pedashenko, H., Pérez-Haase, A., Peterka, T., Prokhorov, V., Rašomavičius, V., Rodríguez-Rojo, M. P., Rodwell, J. S., Rogova, T., Ruprecht, E., Rūsiņa, S., Seidler, G., Šibík, J., Šilc, U., Škvorc, Ž., Sopotlieva, D., Stančić, Z., Svenning, J.-C., Swacha, G., Tsiripidis, I., Turtureanu, P. D., Uğurlu, E., Uogintas, D., Valachovič, M., Vashenyak, Y., Vassilev, K., Venanzoni, R., Virtanen, R., Weekes, L., Willner, W., Wohlgemuth, T. & Yamalov, S. 2016. European Vegetation Archive (EVA): an integrated database of European vegetation plots. Applied Vegetation Science 19/1: 173-180.

Mikula, K.; Ševčovič, D. 2001. Evolution of plane curves driven by a nonlinear function of curvature and anisotropy. SIAM Journal on Applied Mathematics 61, 1473-1501.

Mikula, K.; Ševčovič, D.; Balažovjech, M. 2010. A simple, fast and stabilized flowing finite volume method for solving general curve evolution equations. Communications in Computational Physics 7, 195-211.

Mikula, K., Šibíková, M., Ambroz, M., Kollár, M., Ožvat, A. A., Urbán, J., Jarolímek, I. & Šibík, J. 2021a. NaturaSat—A Software Tool for Identification, Monitoring and Evaluation of Habitats by Remote Sensing Techniques. Remote Sensing 13/17: 3381.

Mikula, K.; Urbán, J.; Kollár, M.; Ambróz, M.; Jarolímek, I.; Šibík, J.; Šibíková, M. 2021b. Semi-automatic segmentation of natura 2000 habitats in sentinel-2 satellite images by evolving open curves. Discrete and Continuous Dynamical Systems - series S, 14, 1033-1046, doi:10.3934/dcdss.2020231.

Mikula, K.; Urbán, J.; Kollár, M.; Ambróz, M.; Jarolímek, I.; Šibík, J.; Šibíková, M. 2021c. An automated segmentation of NATURA 2000 habitats from Sentinel-2 optical data. Discrete and Continuous Dynamical Systems - series S 14, 1017-103.

Šibík, J. 2012. Slovak Vegtation Database. In Vegetation databases for the 21st century, Dengler, J., Oldeland, J., Jansen, F., Chytrý, M., Ewald, J., Finckh, M., Glöckler, F., Lopez-Gonzalez, G., Peet, R.K., Schaminée, J.H.J., Eds.; Biodiversity & Ecology: Hamburg, Germany, Volume 4, pp. 429–429.

Appendix 7. Super-resolution enhancement of Satellite Imagery

Partners involved:

- Jussila Tytti tytti.jussila@syke.fi (Suomen ympäristökeskus - Syke)
- Verbesselt Sebastiaan sebastiaan.verbesselt@inbo.be (Research Institute for Nature and Forest - INBO)
- Stien Heremans stien.heremans@inbo.be (Research Institute for Nature and Forest - INBO)
- Bjärhall Albin albin.bjaerhall@eurac.edu (Eurac Research)
- Skarpman Johanna johanna.skarpman-sundholm@metria.se (Metria)
- Gumaelius Nils nils.gumaelius@metria.se (Metria)
- Moeslund Jesper Erenskjold jesper@ecos.au.dk (Aarhus University)
- Ngo Manh Cuong nm.cuong@ecos.au.dk (Aarhus University)

Background

Sentinel-2 satellite data is already widely used in ecological monitoring, including for grassland and wetland habitat types (e.g., Bartold et al., 2024; Bhatnagar et al., 2020; Close et al., 2021; DeLancey et al., 2022; Dusseux et al., 2022; Ivanova et al., 2023; Li et al., 2021). It is particularly suitable for transnational, long-term biodiversity monitoring thanks to its global coverage, consistent processing, high revisit frequency, long time series, and high availability. However, the coarse spatial resolution of satellite data—ranging from 10 m to 60 m—limits its ability to detect and accurately delineate fine-scale landscape features, thereby also limiting its usability for monitoring certain habitat features or even habitat types. Furthermore, spectral signatures of coarse pixels are more prone to “mixed” signals of reflectance of different ‘ground features’ (e.g. vegetation and water). This problem has long been studied in the remote sensing domain, with techniques such as pansharpening or spectral unmixing algorithms, both of which have their limitations and challenges.

A new field in computer science tries to address these limitations. Deep learning-based super resolution models have emerged as a tool for enhancing the spatial resolution of Sentinel-2 imagery as they can be trained to recognize and reconstruct small features and patterns using the coarse Sentinel-2 data as input (Akhtman, 2022; 2023; Donike et al., 2025 and Wolters et al., 2023). In an ideal scenario, these enhanced-resolution images retain the temporal resolution of Sentinel-2 data while offering improved spatial resolution. This could open for new possibilities in ecological analysis, such as more precise habitat delineation, improved detection of land-management activities like mowing or grazing, and better tracking of short-term natural events including inundations, wildfires, and flooding.

In the Habitat Pilot, we have evaluated two super resolution models—Satlas and S2DR3—through both quantitative performance analyses and qualitative tests focused on their practical utility for ecological monitoring. The goal was to test the performance of the models in different areas, both in terms of pixel fidelity compared to the standard Sentinel-2 data, ability to detect and properly reconstruct fine-resolution details such as line features and small structures in the landscape, and

finally to test super resolution images in combination with habitat quality monitoring spectral indicators to assess the potential of applying the method in ecological monitoring.

Recommendations

We recommend continued testing and evaluation of available super resolution models, with a particular focus on their performance in applied ecological contexts to assess the practical utility of super resolution imagery for ecological monitoring compared to direct usage of Sentinel-2 data. A key-question is whether super resolution should be considered a reliable preprocessing step that enhances Sentinel-2 imagery for ecological analysis, or whether models could instead be trained to interpret Sentinel-2 data directly—potentially bypassing the need for resolution enhancement. Furthermore, it would be valuable to explore the performance of a Satlas model trained specifically on European datasets. Regional training could improve model accuracy by better capturing the spectral and structural characteristics of European landscapes.

The lack of benchmark datasets designed for habitat and ecological monitoring is a serious barrier restricting the comprehensive evaluation of this new technology. Super resolution images can, like other deep learning models, “hallucinate” instead of generating meaningful features. Evaluation of these models is traditionally expressed by pixel-wise metrics (spectral accuracy), perceptual- (spatial accuracy) and model-based metrics. For model based metrics however, we first need performant “remote sensing” models for predicting and monitoring habitats. For example, segmentation models trained on good ground truth data, such as in situ-validated habitat maps, or Contrastive Language-Image Pre-training (CLIP)-scores based on more accurate “land cover / habitat” text labels (instead of general land cover labels from OpenStreetMap (OSM) by Wolters et al., 2023). Ensuring scientific and analytical reliability thus remains a key challenge for operational adoption.

We found that the S2DR3 model looks promising for monitoring dynamic habitats. It provides a 1 x 1 m spatial resolution output for 10 spectral bands of Sentinel-2, in contrast to the Satlas model, which only provides a 2.5 x 2.5 spatial resolution output in RGB. Other bands (the NIR and two SWIR bands) had to be pansharpened, with limited success. Last, the S2DR3 model is a single input super resolution model (SISR: the model uses only one low-resolution image as input to predict a super resolution image) while the Satlas model has originally been trained as a multiple input super resolution model (MISR: the model uses several low-resolution images as an input to predict a super resolution image). This makes the latter model less suitable for monitoring habitats. MISR models tend to reach higher reconstruction accuracy for static objects (like buildings) but its performance decreases with dynamic landscapes/habitats due to sensitivity to input variability and image co-registration.The major drawback of the S2DR3 model is the lack of transparency of the model and the datasets it is trained and validated on by the model developers.

Super resolution modelling is a rapidly evolving field, with continued developments expected in the coming years. It has been successfully implemented in commercial satellite companies as UP42 (together with Nara Space Technology), Maxar Technologies, Airbus SE and recently Planet Labs (Buczkowski, 2023; Davis, 2025). The open-source availability of the Satlas model and the public testing of the Gamma Earth model have drawn attention to the potential in pairing deep learning-based image enhancement with Sentinel-2 data, also encouraging other companies and research groups, e.g., Bhaskar (with OpenCV.ai) (2019), Lim et al. (2017), Malhan (2021), Pereira-Sánchez et al. (2025) and Weber (in TensorFlow) (2019) to train or develop their own super resolution models

and make it open for the community. If super resolution is to be integrated more broadly into transnational biodiversity monitoring efforts across Europe, access to locally trained, open-source models would be highly advantageous. Relying on proprietary solutions—such as S2DR3, despite its current availability for public testing—would likely limit transparency, reproducibility, and long-term sustainability.

Work description

Methods and data

We compared the Satlas model (Wolters et al., 2023), which is the MISR ESRGAN model with the S2DR3 model (Akhtman, 2023), which is a SIRS model. The technical details of both models are given in A7_Table 1

A7_Table 1 Technical details about the two super resolution models.

Author	Wolters, Bastani, & Kembhavi, 2023	Akhtman, 2023
Type	Multi-input super resolution (MISR)	Single-input super resolution (SISR)
Model	ESRGAN variant	Not known. Probably an ESRGAN variant (Akhtman, 2025).
Training dataset	Urban set: S2-NAIP, OSM (in case CLIP loss is added (Radford et al., 2021))	S2 with small samples of 8-bit RGB imaging data exported from Google Earth
Application purpose	Mapping urban areas with high accuracy.	Delineation of agricultural parcel borders (Deeply Resolved Imagery by DigiFarm. (z.d.), Chanev et al., 2025).
Potential purpose	Mapping slow changing habitats and landscapes with high accuracy.	Mapping dynamic habitats and landscapes.
Spatial resolution of product	2.5 meter	1 meter
Spectral resolution of product	3 bands (R, G, B). Other bands were pan-sharpened (by Metria team)	10 bands (B02, B03, B04, B05, B06, B07, B08, B8A, B11, B12) from the MSI of Sentinel-2).

We evaluated both models based on a pixel-wise comparison with low-resolution Sentinel-2 and high-resolution ortho images (and the S2DR3 model with a high resolution satellite image (Pléiades) for a site in Denmark). Pixel-wise comparison is often used in evaluating super resolution images with some variation in which metrics researchers prefer to use in the comparison. The metrics used in this study were R^2, RMSE, PSNR and SSIM (see 'Pixel-wise comparison with reference imagery' for a more detailed description of the metrics) although Wolters, Bastani, & Kembhavi (2023) have argued that traditional metrics, including e.g., PSNR and SSIM are not sensitive enough to so-called "blurred" pixels for a robust performance evaluation. Additionally, both super resolution models have

different spatial resolutions (1 meter versus 2.5 meter spatial resolution) which further complicates their comparison . Therefore, we also evaluated the model output based on human preference (expert evaluation).

Validation sites and habitat types

A7_Table 2 Overview of the field sites and evaluation methods applied for testing the two SR models

Region	Validation site	Habitat types	Model tested	Testing method	Link to github repository / Zenodo
Bolzano	Alpe di Siusi	Grassland	Satlas and S2DR3	Quantitative	https://github.com/albbja/habitat_pilot/tree/main/Super-resolution
	Lago di Caldaro	Wetland	S2DR3	Applied	
Denmark	Molslaboratoriet	Grassland	S2DR3	Quantitative	
Finland	Hirvisuo	Wetland	Satlas and S2DR3	Applied	
Flanders	Webbekomsbroek	Grassland and wetland	Satlas	Quantitative	https://github.com/inbo/BiodiversaHabitatPilot/tree/master/Super%20resolution https://github.com/inbo/BiodiversaHabitatPilot/tree/master/Super%20resolution/output
	Webbekomsbroek	Grassland and wetland	S2DR3	Quantitative/Applied	https://github.com/inbo/BiodiversaHabitatPilot/tree/master/Super%20resolution https://github.com/inbo/BiodiversaHabitatPilot/tree/master/Super%20resolution/output For the applied case, see the inundation subtask report.
	Schulensmeer	Grassland and wetland	S2DR3	Applied	
	Kloosterbeemden	Grasslands and wetlands	S2DR3	Applied	
Sweden	Stockholm	Urban	Satlas	Explorative/Quantitative	
	Brösarp	Grassland and cropland	Satlas	Explorative/Quantitative	
	Vemdalen	Urban and forest	Satlas	Explorative/Quantitative	

Super resolution modelling

Satlas is a MISR model developed by the Allen Institute for AI as an adaptation of the ESRGAN model described by Wang et al. (2018). Satlas is available on Github both for training and as a pre-trained model. The pre-trained model, which was used for the evaluations in the Habitat Pilot, has been trained on downscaled aerial images around cities in the USA from the USGS National Agriculture Imagery Program (2.5 meter resolution) paired with matching Sentinel-2 data (10 meter resolution). The evaluated super-resolution images were developed using eight L1C Sentinel-2 tiles of 32x32 pixels (10 meter resolution) to create a tile of 128x128 pixels with enhanced (2.5 meter) resolution. Tiles were then combined by overlapping each tile (with an overlap of 0.5 times the size of the tile) and a weight map preferring the central over the peripheral pixel values to avoid sharp edges. More information about the process of generating the Satlas images for evaluation in the Habitat Pilot, including some additional conclusions can be found in a report by Gumaelius et al. (2025).

S2DR3 is a SISR model developed by Gamma Earth. The high resolution dataset used for training the model, and the model type is unknown, although Akhtman (2022) describes the underlying model architecture as “not dissimilar” to that of ESRGAN. S2DR3 is only available as a pre-trained model, accessible for public testing through a Google Colab script (S2DR3T-infer-20240430.ipynb). Through the script, users can download 4x4 km² tiles of L2A Sentinel-2 data for a specified date and area together with super-resolved images of all 10 Sentinel-2 spectral bands with ten times the original spatial resolution. More information about the model can be found in an article written by Akhtman (2022) from Gamma Earth.

Pixel-wise comparison with reference imagery:

For the quantitative pixel-wise comparisons, the evaluation method described by Akhtman (2022) was recreated in R. According to the method, super-resolved images of a Sentinel-2 scene are histogram-matched and re-scaled to the original Sentinel-2 resolution before the pixel values are scatter plotter and R², RMSE, PSNR and SSIM are calculated. The same method was also applied to evaluate the RGB bands of the super-resoluted images in their full resolution. This was done by using an orthophoto scaled down to the resolution of the super-resolution images as reference image. Additionally, we compared the super-resolved pixels to a same-day Pléiades 0.5 m resolution image in the test area in Denmark. Before comparison, this image was downsampled to 1 m resolution to exactly match the super-resolution image. This test is described in detail in Erenskjold Moeslund & Manh Cuong Ngo (2025).

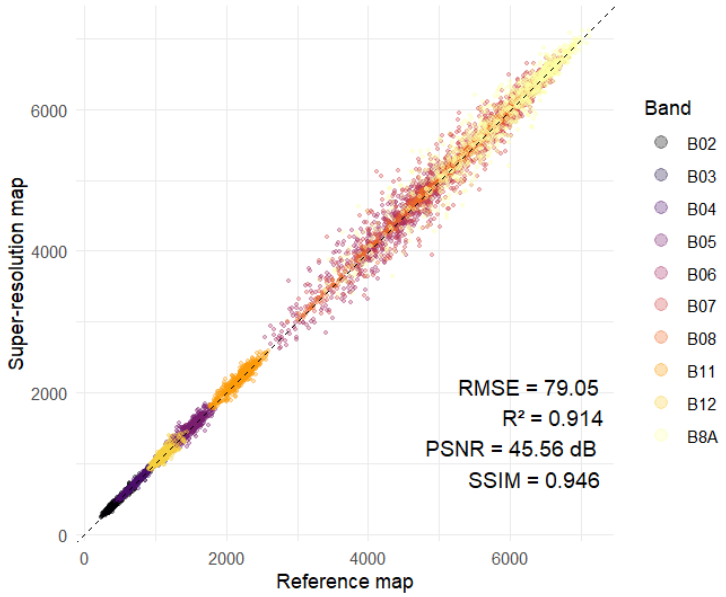
A7_Table 3 Descriptions of the metrics used for the pixel-wise comparison of the super resolution models

Metric	Description	Calculation
Coefficient of Determination (R^2) of the 1:1 relation.	R^2 ranges from -inf–1, with 1 being a perfect fit. Negative values indicate that the model is arbitrarily worse.	$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$ $SS_{res} = \sum_i (y_{true,i} - y_{pred,i})^2$

		$SS_{tot} = \sum_i \left(y_{true,i} - \underline{y}_{true}\right)^2$
Root Mean Square Error (RMSE)	RMSE calculates the pixel-wise error between the super resolution image and the reference image with a lower number indicating more resemblance in the pixel intensity between the two images	$RMSE = \sqrt{\frac{1}{n}\sum_i \left(y_{true,i} - y_{pred,i}\right)^2}$
Peak Signal-to-Noise Ratio (PSNR)	PNSR is widely used for evaluating the quality of reconstructed or corrupted images and videos. It is based on the MSE between the super resolution and reference image, comparing the pixel values with their maximum pixel value. PSNR is expressed in dB and a higher value generally indicates a better image quality, although the correlation with human perception of an image has been found to be low.	$\mathrm{PSNR} = 20\log_{10}(\mathrm{MAX}_{\mathrm{true}}) - 10\log_{10}(\mathrm{MSE})$ $MAX_{true} = 1000$ (the surface-reflectance scale used in Sentinel-2 L2A data)
Structural Similarity Index Measure (SSIM)	SSIM evaluates the structural information, luminance and contrast information by comparing mean, standard deviation and cross-covariance of pixel intensities within local windows in the super resolution and reference image. SSIM is more computationally demanding than the other metrics, but should correlate better with human perception of an image. SSIM ranges from 0–1 with higher values indicating a higher structural similarity.	SSIM was calculated using the R-package SSIMmap (Ha, 2023)

Results

Pixel-wise comparison with reference imagery

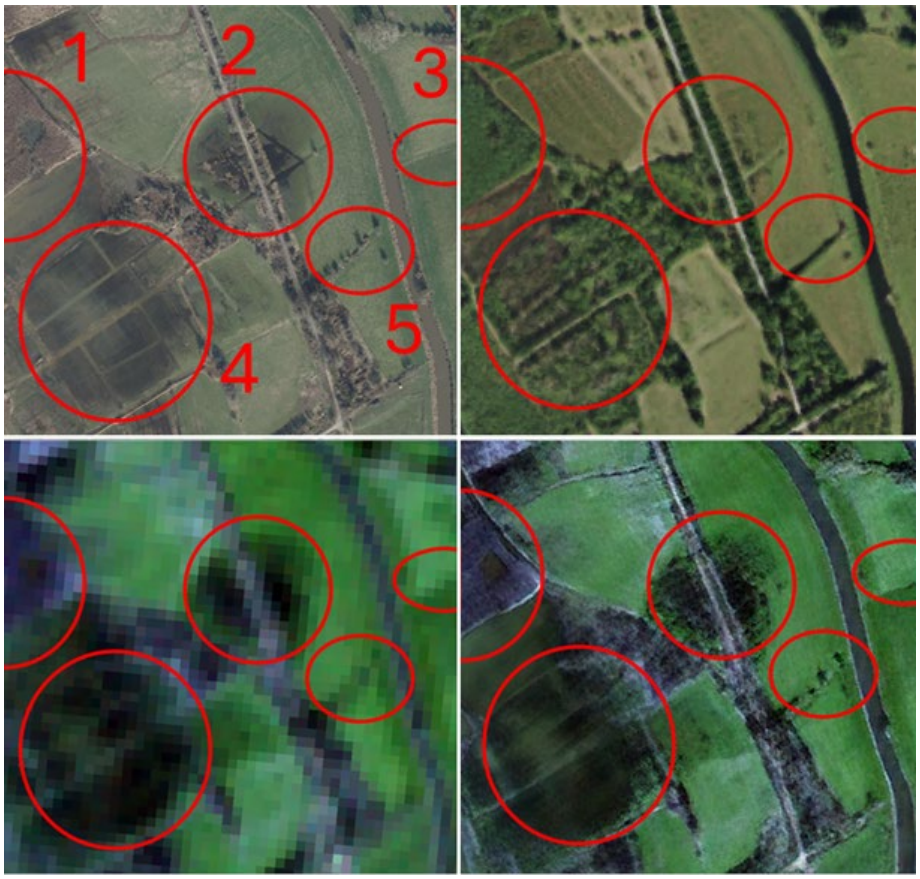


A7_Fig 1 Example of the pixel-wise SR model comparison with reference imagery. The figure shows the comparison of a histogram-matched and re-scaled S2DR3 image with a Sentinel-2 scene.

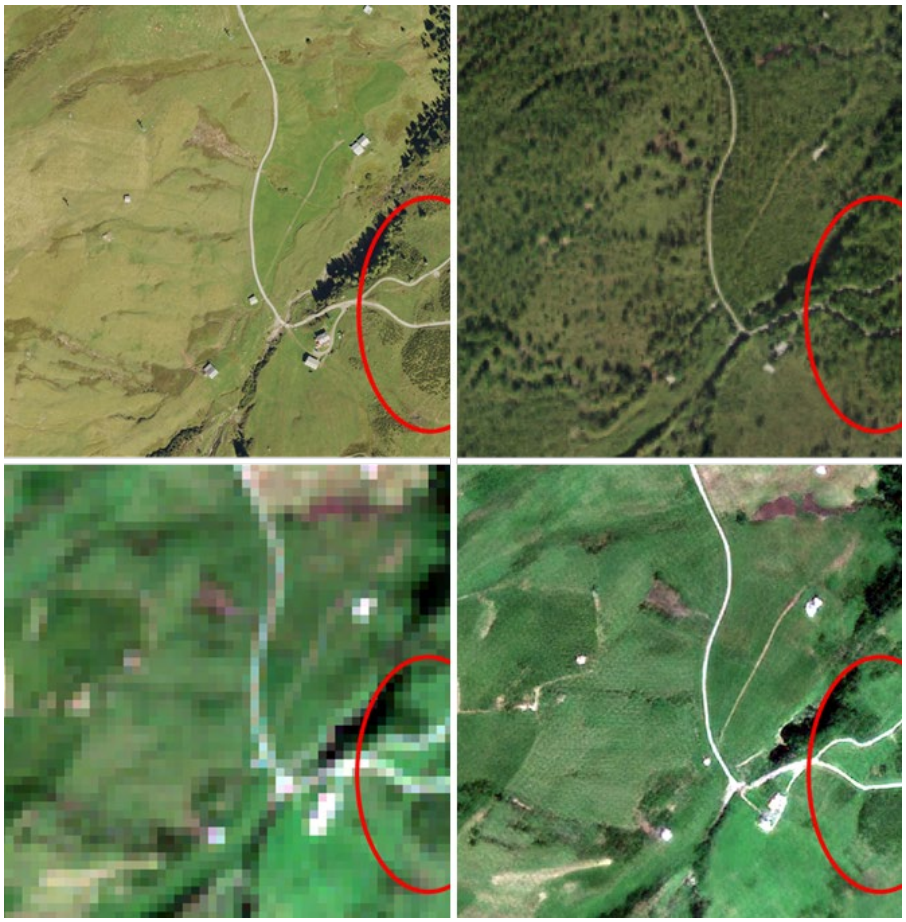
A7_Table 4 Summary of pixel-wise comparison metrics calculated by the partners. SR images in analyses by Bolzano and Flanders were histogram matched with the reference image, and higher-resolution image of each pair was re-scaled to match the other image.

Partner (ref. sensor, SR model)	R^2	RMSE	PSNR	SSIM
Bolzano (S2, Satlas)	-0.767	308.16	34.66	0.472
Bolzano (S2, S2DR3)	0.914	79.05	45.56	0.946
Bolzano (ortho image, Satlas)	-0.694	28.6	51.25	0.142
Bolzano (ortho image, S2DR3)	0.042	23.5	53.12	0.335
Flanders (S2, Satlas)	0.222	406.88	29.09	0.383
Flanders (S2, S2DR3)	0.948	115.08	40.07	0.887
Flanders (ortho image, Satlas)	-0.185	1212.95	18.39	0.089
Flanders (ortho image, S2DR3)	0.412	918.4	20.78	0.157
Denmark (Pléiades, S2DR3)	-3.8–(-5.0)	28–46	14–19	0.6–0.7

Human preference (expert evaluation)

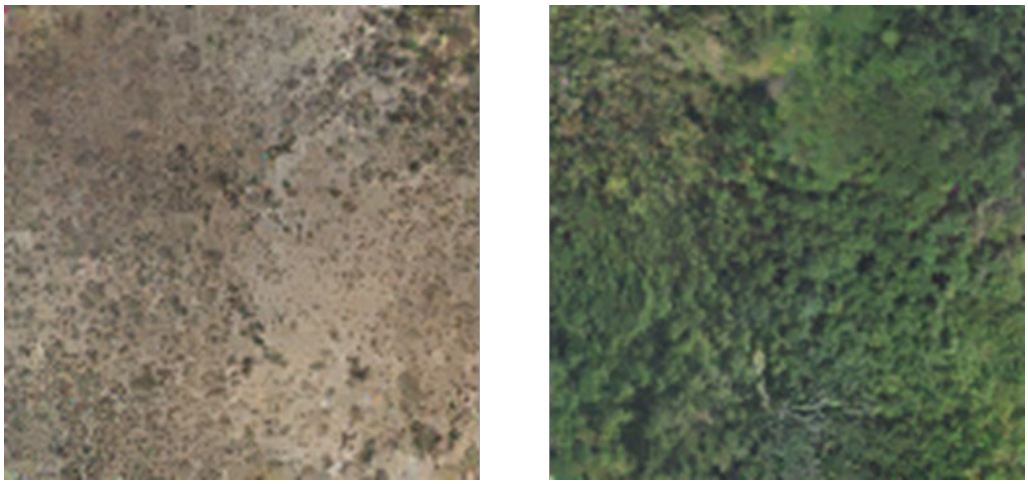


A7_Fig 2 Example demonstrating the distinguishability of landscape features including a reedbed (1), an inundated forest area (2), a river (3), an inundated open area (4), and a tree line (5) visualized in an ortho photo (t.l.), Sentinel-2 scene (b.l.), Satlas image (t.r.), and S2DR3 image (b.r.)



A7_Fig 3 Example demonstrating tree-like 'hallucinations' of the Satlas model (t.r.) making e.g., the detection of Rhododendron shrubs (circled) more difficult than in the S2DR3 image (b.r.). Orthophoto (t.l.) and Sentinel-2 scene (b.l.) included for reference.

When we compared the Satlas model with the S2DR3 model, most of the researchers within the group had the impression that the model tends to "overrepresent" features from the NAIP ortho images. This may be caused by overfitting of the model to this dataset, and that the extrapolation to European sites results in too high uncertainty (although some sites in France were included in the test dataset). An additional experiment was done with the Satlas ESRGAN model to test this "phenomenon". On the Github page, there is an ESRGAN variant that works with one Sentinel-2 input image (this model was not published in the paper of Wolters, Bastani, & Kembhavi, 2023). By feeding the pre-trained model with a completely black or white (for the three channels - RGB) input image (without any real information), we were able to systematically produce the output displayed in A7_Fig 4. This process was tested multiple times to assess the potential uncertainties in the process.



A7_Fig 4 Systematic output for a complete black (left) or white (right) 32 x 32 pixel input image of Sentinel 2 (10m x 10m spatial resolution). The model "hallucinates" a standard output of in RGB of 128 x 128 pixels (2.5m x 2.5m spatial resolution).

The examined output resembles bare land with shrubs (A7_Fig 4- left) and more densely vegetated (A7_Fig 4- right) area that is quite similar to NAIP images of Arizona (e.g. https://labs.mapbox.com/labs/naip/#20.02/34.0054669/-113.1829979 and https://labs.mapbox.com/labs/naip/#19.01/34.5510688/-111.0079828). We acknowledge here that this sort of comparison, as well as and other qualitative model comparisons, should ultimately be validated by targeted in situ site surveys.

Finally, in addition to the pixel-wise comparison and the human preference evaluations of the two models, images generated by the S2DR3 model were also combined with a model for classifying inundation in open wetlands using Sentinel-2 data developed by Jussila et al. (2024). These tests, carried out in sites in Flanders, Finland and Bolzano are described more in detail in Appendix 2: Inundation mapping.

Discussion

Comparing the performance of super-resolution models for remote sensing data is challenging due to differences in input data, output resolution, and reference imagery. The lack of good benchmarks and variation in model performance metrics also limit the conclusions that can be drawn from quantitative analyses.

Additionally, there are inherent differences in the design and the intended use of different models that limit their comparability. For example, the images generated with the Satlas model were trained with Sentinel-2 scenes from up to a two-month period, compared to the single Sentinel-2 scene used to generate the images of the S2DR3 model. This margin of error of up to two months' time, compared to the reference image, can have a large impact on the spectral features of the landscape, especially in dynamic habitats like grasslands and wetlands. This will favor comparisons between S2DR3 images and reference data from the same date over comparisons between Satlas images and reference data from a date in the corresponding time period.

Furthermore, the resolution of the super-resolved images differs significantly: Satlas produces RGB outputs at 2.5 × 2.5 m, whereas S2DR3 generates 1 × 1 m outputs across 10 spectral bands. This

results in a large difference in total number of pixels for a given area between the two models, which also affects the calculated metrics.

Another shortcoming is the limited access to high-resolution referential datasets that allows for consistent transnational evaluation of super resolution images at full resolution. When super resolution images are re-scaled for comparison with lower-resolution Sentinel-2 data, their full potential is not assessed. Conversely, comparisons with orthophotos introduce inconsistencies across countries due to the use of different sensors in the creation of orthophotos. The use of high-resolution satellite data, such as PlanetLabs or Pléiades, could mitigate these inconsistencies. However, acquiring such datasets come at a cost which could make large-scale evaluations expensive. Moreover, the evaluations performed in Denmark using a Pléiades satellite scene showed unsatisfactory results for R^2, RMSE and PSNR, probably in part due to the differences in the sensors and the atmospheric correction methods used for Sentinel 2 and Pléiades.

Finally, the two models evaluated are developed for different purposes: the MISR Satlas model is developed for leveraging several Sentinel-2 scenes to give a cloud-free, high-resolution image output, and tests in Flanders and Sweden showed that the model greatly benefitted from using eight Sentinel-2 scenes as model input, compared to only using one. However, many potential use cases offered by higher-resolution Sentinel-2 imagery, e.g., tracking inundation or mowing events, are lost if the temporal resolution is 1-2 months instead of the regular 5-6-day revisit frequency of Sentinel 2. The Satlas model also showed tendencies to 'hallucinate' as well as to change the general aspect of landscapes, e.g., by changing both the coloration and texture of landscape most likely in part due to the difference between the European grassland and wetland habitat evaluated in the Habitat Pilot compared to the North American urban landscapes used to train the model. It would therefore be interesting to see if the performance of the Satlas model could be improved by training it using local European data including more natural environments, and if this could also improve the model hallucinations.

Despite only needing one Sentinel-2 image as input, we found the S2DR3 image to be promising for application in monitoring dynamic habitats. Applied tests using the Jussila model for mapping the dynamics of wetland water surfaces (see Appendix 2: Inundation mapping) showed a maintained pixel classification performance using S2DR3 image as using Sentinel-2 data, but with the added benefit of higher spatial resolution. These results were in line with a study by Chanev et al. (2025) comparing S2DR3 imagery with Sentinel-2 data for classifying winter crops. However, the lack of transparency regarding model training and validation and the limited access to model testing remain significant drawbacks.

Accurately representing fine-resolution structures and maintaining spectral fidelity with Sentinel-2 data remain the two main challenges for super-resolution models of Sentinel-2 imagery. In our tests, the two models performed better at creating high-resolution images for visual evaluation than in the pixel-wise comparisons of the SR-model output with high resolution ortho images. Simultaneously, applied tests using S2DR3 images to map inundation show some potential in using super-resolution images also for reflectance-based analysis. This highlights the need both for continued testing of super resolution models, using accurate global benchmarks, and methods and metrics capturing both spatial accuracy, spectral accuracy, and the human perception of the imagery.

References

Akhtman, Y. (2022). Sentinel-2 Deep Resolution 2.0. S2DR2: Effective 12-Band 10x Single…. https://medium.com/@ya_71389/sentinel-2-deep-resolution-2-0-c3d530d9bdf8

Akhtman, Y. (2023). Sentinel-2 Deep Resolution 3.0. Medium. https://medium.com/@ya_71389/sentinel-2-deep-resolution-3-0-c71a601a2253

Bartold, M., Kluczek, M., Wróblewski, K., Dąbrowska-Zielińska, K., Goliński, P., & Golińska, B. (2024). Mapping management intensity types in grasslands with synergistic use of Sentinel-1 and Sentinel-2 satellite images. Scientific Reports, 14(1), 32066. https://doi.org/10.1038/s41598-024-83699-4

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," 2nd NTIRE: New Trends in Image Restoration and Enhancement workshop and challenge on image super-resolution in conjunction with CVPR 2017.

Bhaskar, Abhinav. 'From Blurry to Brilliant: Upscaling Satellite Images Using OpenCV DNN'. Medium, 31 januari 2025. https://medium.com/@animagun/from-blurry-to-brilliant-upscaling-satellite-images-using-opencv-dnn-efe5c6be6618

Bhatnagar, S., Gill, L., Regan, S., Naughton, O., Johnston, P., Waldren, S., & Ghosh, B. (2020). MAPPING VEGETATION COMMUNITIES INSIDE WETLANDS USING SENTINEL-2 IMAGERY IN IRELAND. International Journal of Applied Earth Observation and Geoinformation, 88, 102083. https://doi.org/10.1016/j.jag.2020.102083

Buczkowski, Aleks. 'Unlocking a Clearer World: An Overview of Planet SuperRes'. 16 mei 2023. https://www.planet.com/pulse/unlocking-a-clearer-world-an-overview-planet-superres/

Chanev, M., Kamenova, I., Dimitrov, P., & Filchev, L. (2025). Evaluation of Sentinel-2 Deep Resolution 3.0 Data for Winter Crop Identification and Organic Barley Yield Prediction. Remote Sensing, 17(6), 957. https://doi.org/10.3390/rs17060957

Chanev, M., Kamenova, I., Dimitrov, P., & Filchev, L. (2025). Evaluation of Sentinel-2 Deep Resolution 3.0 Data for Winter Crop Identification and Organic Barley Yield Prediction. Remote Sensing, 17(6), 957. https://doi.org/10.3390/rs17060957

Close, O., Petit, S., Beaumont, B., & Hallot, E. (2021). Evaluating the Potentiality of Sentinel-2 for Change Detection Analysis Associated to LULUCF in Wallonia, Belgium. Land, 10(1), 55. https://doi.org/10.3390/land10010055

Davis, Justin. 'Enhancing Satellite Imagery Readability with Super-Resolution Machine Learning Models'. Geoawesome, 2 oktober 2025. https://geoawesome.com/eo-hub/enhancing-satellite-imagery-readability-with-super-resolution-machine-learning-models/

Deeply Resolved Imagery by DigiFarm. (z.d.). Accessed on 28th of October 2025, from https://digifarm.io/products/dr-imagery

DeLancey, E. R., Czekajlo, A., Boychuk, L., Gregory, F., Amani, M., Brisco, B., Kariyeva, J., & Hird, J. N. (2022). Creating a Detailed Wetland Inventory with Sentinel-2 Time-Series Data and Google Earth Engine in the Prairie Pothole Region of Canada. Remote Sensing, 14(14), 3401. https://doi.org/10.3390/rs14143401

Donike, S., Aybar, C., Gómez-Chova, L., & Kalaitzis, F. (2025). Trustworthy Super-Resolution of Multispectral Sentinel-2 Imagery With Latent Diffusion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, PP, 1-14. https://doi.org/10.1109/JSTARS.2025.3542220

Dusseux, P., Guyet, T., Pattier, P., Barbier, V., & Nicolas, H. (2022). Monitoring of grassland productivity using Sentinel-2 remote sensing data. International Journal of Applied Earth Observation and Geoinformation, 111, 102843. https://doi.org/10.1016/j.jag.2022.102843

Erenskjold Moeslund, J. & Cuong, N. (2025) Super resolution comparison summary. Unpublished. Available upon request from: jesper@ecos.au.dk

Gumaelius, N., Skarpman Sundholm, J. & Eriksson, T. (2025) Biodiversa Super resolution Test. Unpublished. Available upon request from: camilla.jonsson@naturvardsverket.se

Ha, H. (2024). Hailyee-Ha/SSIMmap [R]. https://github.com/Hailyee-Ha/SSIMmap (Original work published 2023)

Jussila, Tytti, Risto K. Heikkinen, Saku Anttila, e.a. 'Quantifying Wetness Variability in Aapa Mires with Sentinel-2: Towards Improved Monitoring of an EU Priority Habitat'. Remote Sensing in Ecology and Conservation 10, nr. 2 (2024): 172-87. https://doi.org/10.1002/rse2.363

Ivanova, I., Spasova, T., & Stankova, N. (2023). Using Sentinel-2 data for efficient monitoring and modeling of wetland protected areas. Ninth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2023), 12786, 658–665. https://doi.org/10.1117/12.2681790

Jussila, Tytti, Risto K. Heikkinen, Saku Anttila, e.a. 'Quantifying Wetness Variability in Aapa Mires with Sentinel-2: Towards Improved Monitoring of an EU Priority Habitat'. Remote Sensing in Ecology and Conservation 10, nr. 2 (2024): 172-87. https://doi.org/10.1002/rse2.363

Kowaleczko, Pawel, Tomasz Tarasiewicz, Maciej Ziaja, e.a. 'A Real-World Benchmark for Sentinel-2 Multi-Image Super-Resolution'. Scientific Data 10, nr. 1 (2023): 644. https://doi.org/10.1038/s41597-023-02538-9.

Li, C., Zhou, L., & Xu, W. (2021). Estimating Aboveground Biomass Using Sentinel-2 MSI Data and Ensemble Algorithms for Grassland in the Shengjin Lake Wetland, China. Remote Sensing, 13(8), 1595. https://doi.org/10.3390/rs13081595

Liu, Denghui, Lin Zhong, Haiyang Wu, Songyang Li, en Yida Li. 'Remote Sensing Image Super-Resolution Reconstruction by Fusing Multi-Scale Receptive Fields and Hybrid Transformer'. Scientific Reports 15, nr. 1 (2025): 2140. https://doi.org/10.1038/s41598-025-86446-5

Malhan, Kazumi. kmalhan/SatelliteSR. Jupyter Notebook. 17 april 2021, released 8 juli 2025. https://github.com/kmalhan/SatelliteSR

Pereira-Sánchez, Ivan, Daniel Torres, Francesc Alcover, e.a. 'Super-Resolution of Sentinel-2 Images Using a Geometry-Guided Back-Projection Network with Self-Attention'. arXiv:2508.04729. Preprint, arXiv, 5 augustus 2025. https://doi.org/10.48550/arXiv.2508.04729

Radford, Alec, Jong Wook Kim, Chris Hallacy, e.a. 'Learning Transferable Visual Models From Natural Language Supervision'. arXiv:2103.00020. Preprint, arXiv, 26 februari 2021. https://doi.org/10.48550/arXiv.2103.00020

Weber, Xavier. Saafke/EDSR_Tensorflow. Python. 8 juli 2019, accessed on the 6 of November 2025. https://github.com/Saafke/EDSR_Tensorflow

Wolters, P., Bastani, F., & Kembhavi, A. (2023). Zooming Out on Zooming In: Advancing Super-Resolution for Remote Sensing (No. arXiv:2311.18082). arXiv. https://doi.org/10.48550/arXiv.2311.18082