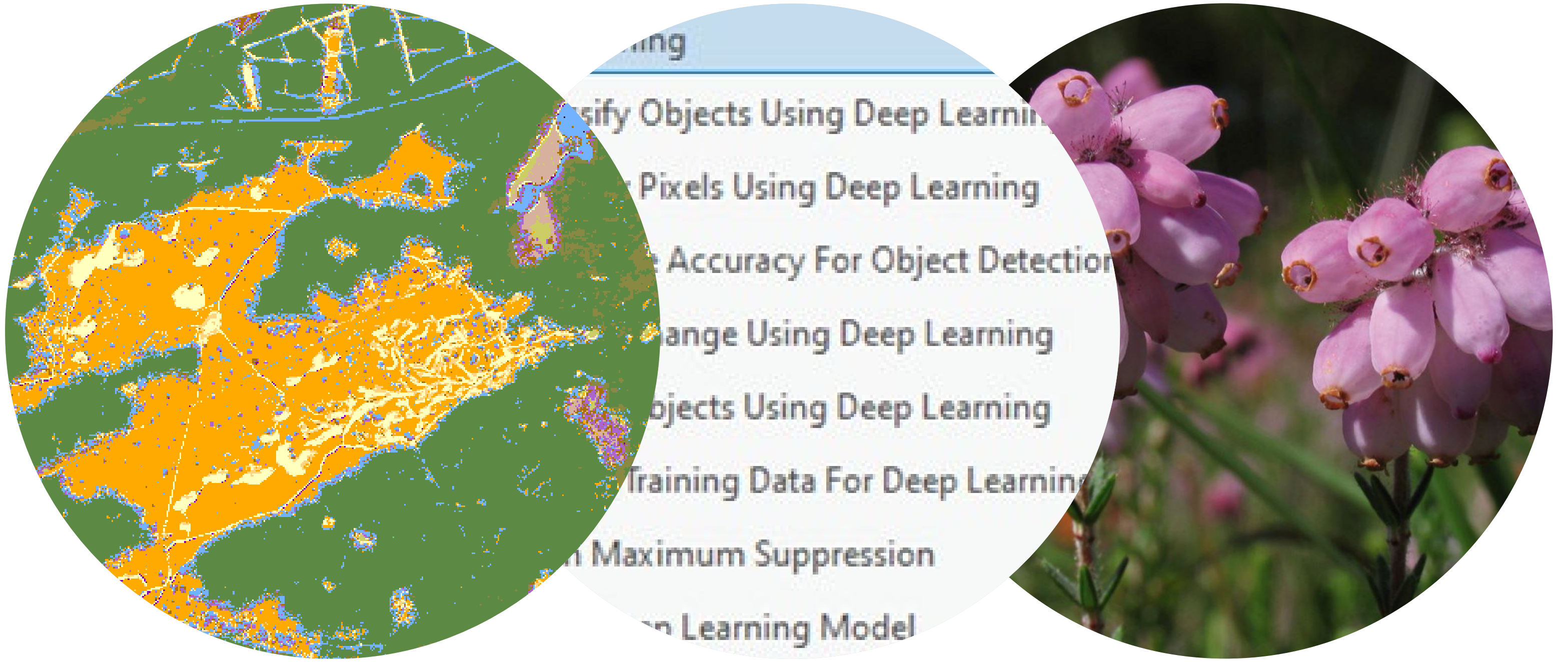
Modelling and Mapping Habitats at European and Regional Scale using AI/ML techniques

EO4EA 2022 Workshop on Earth Observation For Ecosystem Accounting, 29 of November 2022

Sander Mucher[1]; Stephan Hennekens[1]; Bruno Smets[2]; Sara Simoussi[3]; Henk Kramer[1]; Rob Knapen[1]; Marcel Buchhorn[2]; Wilfried Thuiller[3]; Kristof Vantricht[2]; Stan Los[1]

1 Wageningen University and Research, Netherlands; 2 VITO, Belgium; 3 CNRS, France



WAGENINGEN UNIVERSITY & RESEARCH

vito remote sensing

cnrs

Background

- Background is the latest assessment by the EEA (The European environment – state and outlook 2020) shows that Europe's biodiversity continues to decline at an alarming rate, with most protected species and habitats found not to have a good conservation status.
- Much more effort is needed to reverse current trends and to ensure resilient and healthy nature.
- The EU's biodiversity strategy for 2030 is a comprehensive, ambitious and long-term plan to protect nature and reverse the degradation of ecosystems.

Background & objectives

- Artificial Intelligence (AI) techniques, such as Machine Learning (ML) and Neural Networks (NN) or Deep Learning (DL) methods could enable an improved monitoring of biodiversity and ecosystems with satellite based high-resolution datasets such as Copernicus High Resolution Vegetation Phenology Product (HR-VPP) to better support European policy making.
- So, understanding where habitats occur across Europe is a crucial element for understanding biodiversity conservation and taking specific actions
- Our overall objective to exploit AI / deep learning classification methods for habitat mapping WENR wants

Habitat mapping strategy

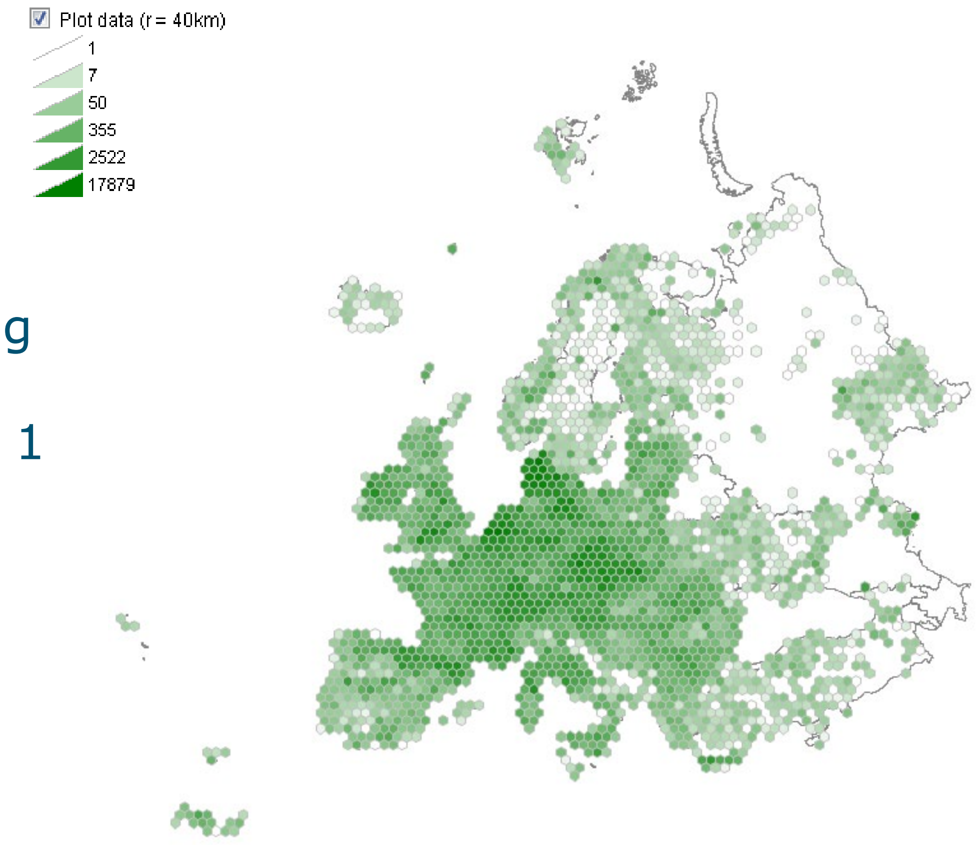
Two approaches:

1. **European** habitat suitability modelling at 100 meter resolution by using RS-enabled EBVs and other bioclimatic layers as predictors in BIOMOD2 (e.g. MAXENT, Maximum Entropy) model trained by exploiting *in situ* vegetation plot data from the European Vegetation Archive (EVA, http://euroveg.org/eva-database)
2. **Regional** habitat mapping using deep learning techniques at 10 or 20 meter resolution

In both approaches training data from the EVA database plays a central role

Method 1 European habitat modelling

- Input for the modelling are 1,2M vegetation plot observations (derived from the European Vegetation Archive (EVA database) covering ~203 EUNIS habitats.
- The Maxent model for each habitat type is executed using a selection of 22 predictors (comprising 5 climate parameters, 7 soil, 2 terrain parameters, 7 RS-EBVs and 1 topography parameter).
- For the habitat modelling open source software Maxent version 3.4.1 is used, by applying a machine-learning technique called Maximum Entropy Modelling.
- We did run MAXENT model to create European habitat suitability maps at 100 meter resolution for most EUNIS habitat types at level 3 (203 EUNIS classes)



WAGENINGEN UNIVERSITY & RESEARCH

European Predictors used

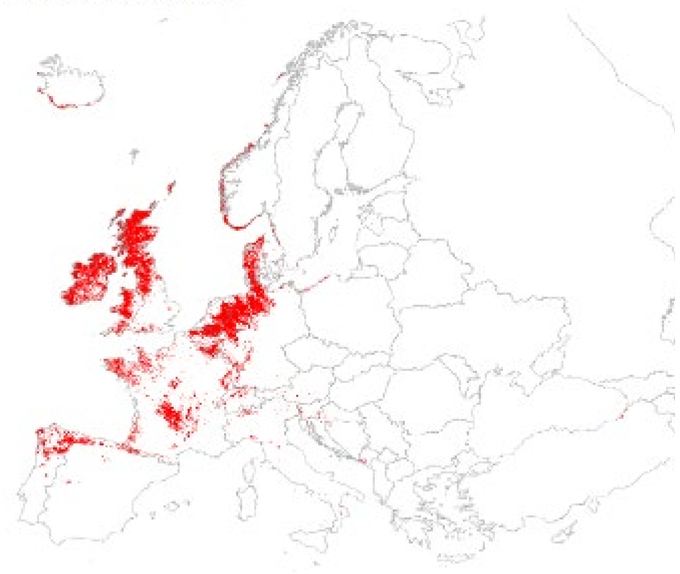
Group	Predictor description	Nr
Climate	Annual precipitation (mm yr-1)	1
	Growing degree days heat sum above 5°C (gdd5)	2
	Accumulated precipitation amount on growing season days TREELIM (gsp)	3
	Mean temperature of the growing season TREELIM (gst)	4
	Snow covered days (scd)	5
Elevation	EU DEM	6
	EU DEM slope	7
HR-VPP	VPP - Season amplitude given by MAXV-MINV	8
	VPP - Length of season (number of days between start and end)	9
	VPP - Slope of the green-up season (*PP I* × *day*-1)	10
	VPP - PPI at the day of maximum-of-season	11
Inundation	Inundation - occurrence	12
Land cover	Corine Land Cover	13
	World cover	14
Soil	Soil - bulk density	15
	Soil - cation exchange capacity	16
	Soil - course fractions	17
	Soil - clay fraction	18
	Soil - pH	19
	Soil - sand fraction	20
	Soil - organic carbon	21
Topography	Distance to inland water	22

Example result Maxent for S41 Wet heath

In total 203 European habitat suitability maps for most EUNIS habitat types

Distribution data (in-situ)

Thresholded suitability map

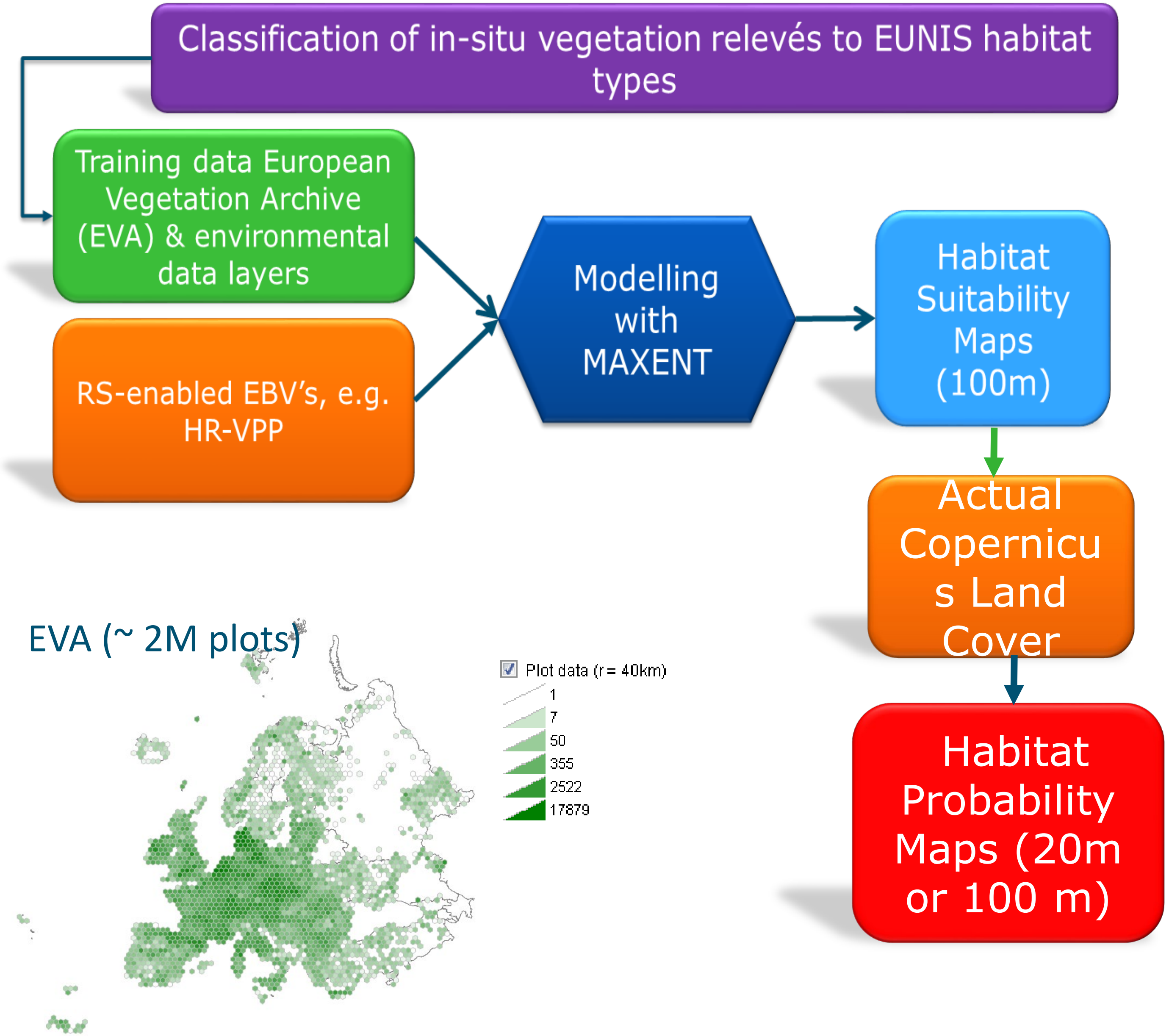


Statistics from Maxent modelling

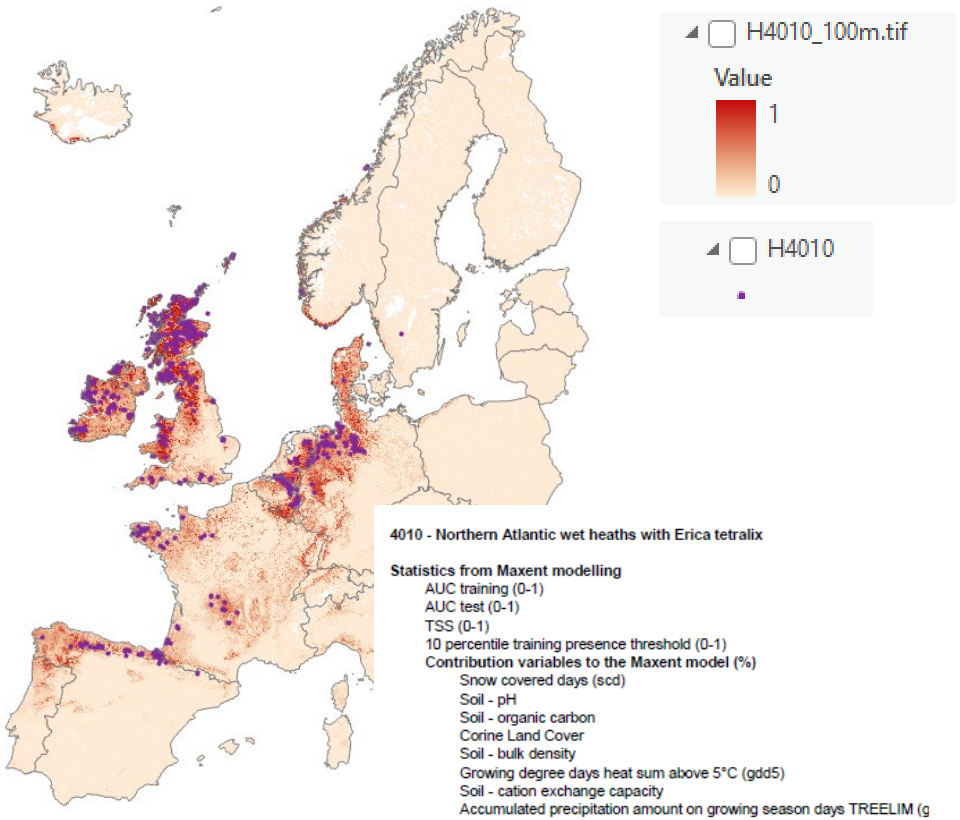
AUC training (0-1)	0.8679
AUC test (0-1)	0.8384
10 percentile training presence threshold (0-1)	0.2797
Contribution variables to the Maxent model (%)	
Climate - Snow covered days (scd)	30.63
Soil - pH	29.7015
Corine Land Cover 2018	10.8223
Climate - Accumulated precipitation amount on growing s	8.8832
EU DEM	7.2818
Climate - Growing degree days heat sum above 5°C (gdc	4.3854
Soil - clay fraction	3.2767
Soil - course fractions	1.372
Soil - bulk density	1.0599
EU DEM Slope	0.9925
Soil - organic carbon	0.5874
HR-VPP - PPI at the day of maximum-of-season	0.4623
Climate - Annual precipitation (mm yr-1)	0.1383
Soil - sand fraction	0.0971
Climate - Mean temperature of the growing season TREE	0.0879
HR-VPP - Slope of the greenup season (PP I × day-1)	0.0831
Soil - cation exchange capacity	0.0458
HR-VPP - Length of season (number of days between sta	0.0426
Population density 2018	0.0418
HR-VPP - Season amplitude given by MAXV-MINV	0.0065
Distance to inland water	0.002
Inundation - occurrence	0

WAGENINGEN UNIVERSITY & RESEARCH

Flowchart European habitat modelling



Suitability map H4010: Northern Atlantic wet heaths with *Erica tetralix & distribution data (training)*



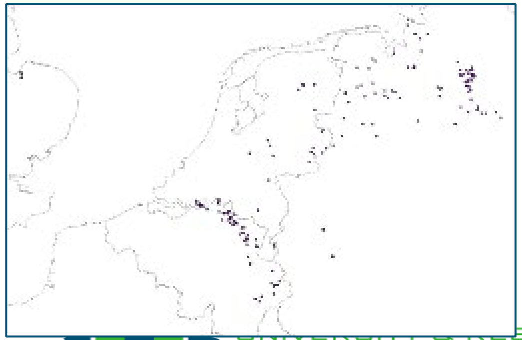
4010 - Northern Atlantic wet heaths with Erica tetralix

Statistics from Maxent modelling	
AUC training (0-1)	0.9018
AUC test (0-1)	0.9006
TSS (0-1)	0.8455
10 percentile training presence threshold (0-1)	0.3927
Contribution variables to the Maxent model (%)	
Snow covered days (scd)	39.0162
Soil - pH	27.9856
Soil - organic carbon	15.1298
Corine Land Cover	5.6031
Soil - bulk density	3.5019
Growing degree days heat sum above 5°C (gdd5)	2.2464
Soil - cation exchange capacity	1.7273
Accumulated precipitation amount on growing season days TREELIM (g	1.7192
HR-VPP - Length of season (number of days between start and end)	0.6452
Annual precipitation (mm yr-1)	0.5998
EU DEM Slope	0.5858
EU DEM	0.445
Soil - clay fraction	0.2858
Soil - sand fraction	0.1996
Population density	0.1179
HR-VPP - Slope of the greenup season (PP I × day-1)	0.0682
Soil - course fractions	0.0491
Mean temperature of the growing season TREELIM (gst)	0.0258
Distance to inland water	0.0182
HR-VPP - Season amplitude given by MAXV-MINV	0.017
HR-VPP - PPI at the day of maximum-of-season	0.013
Inundation - occurrrence	0

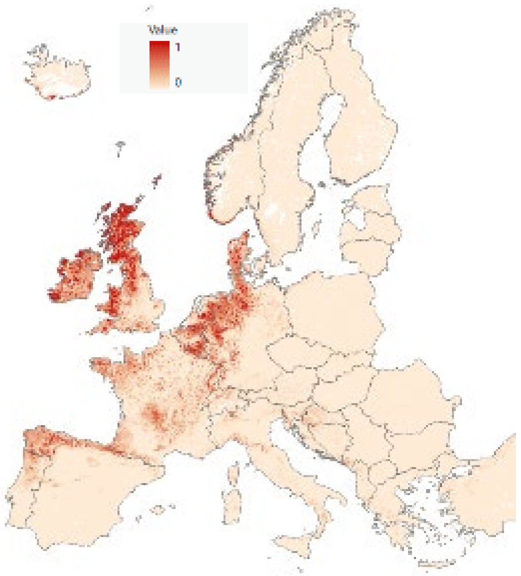
Example European habitat modelling: S41 Wet heath

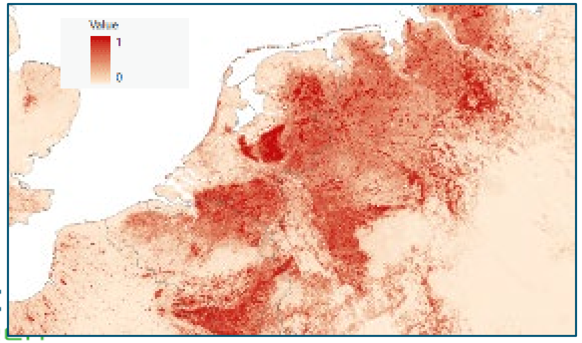
Distribution data

from European Vegetation Archive (EVA)

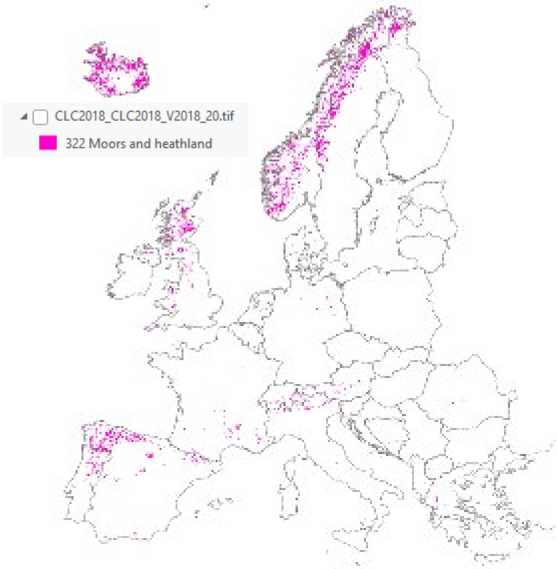


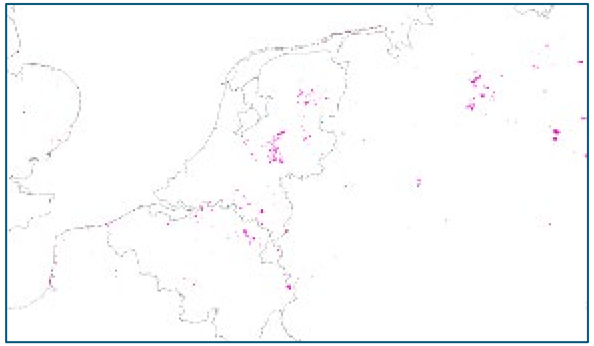
Habitat suitability map



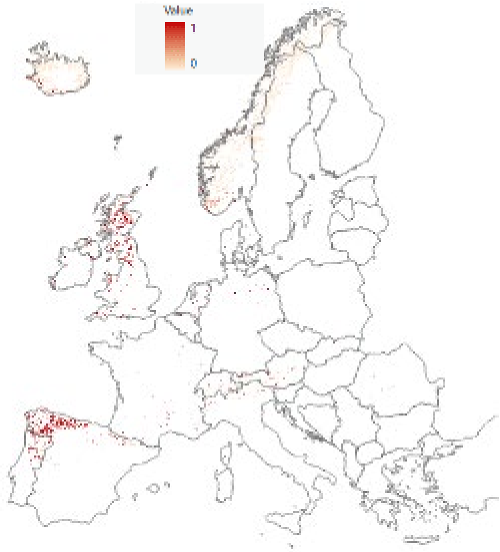


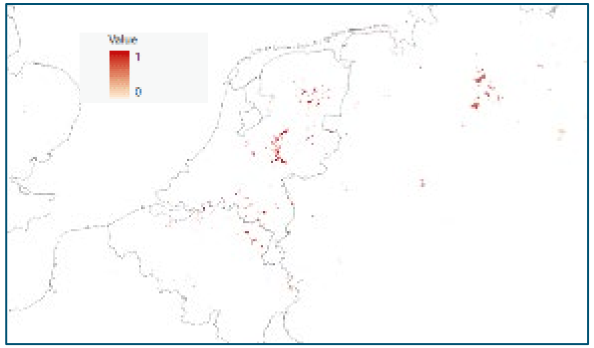
Land cover



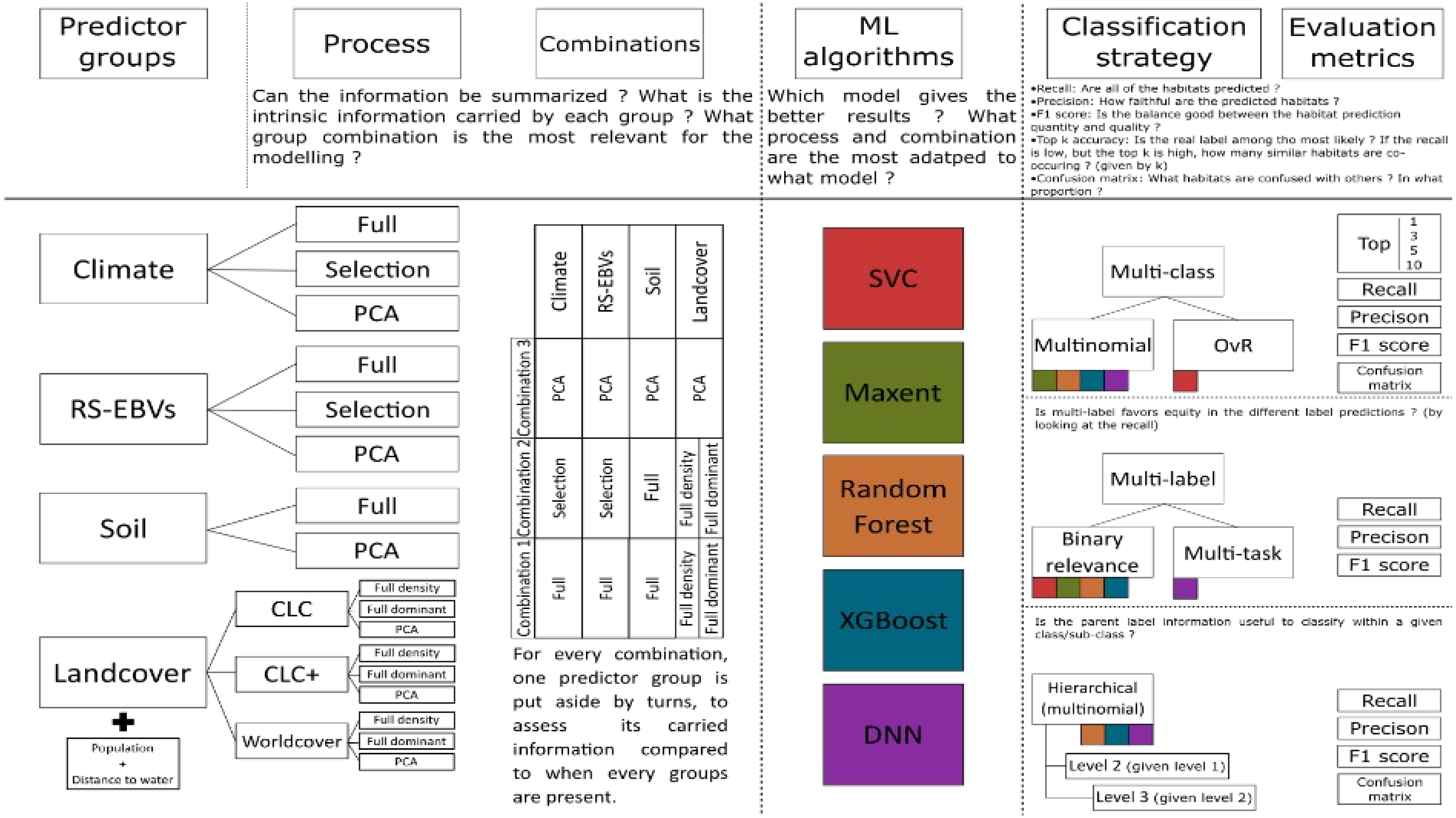


Habitat probability map

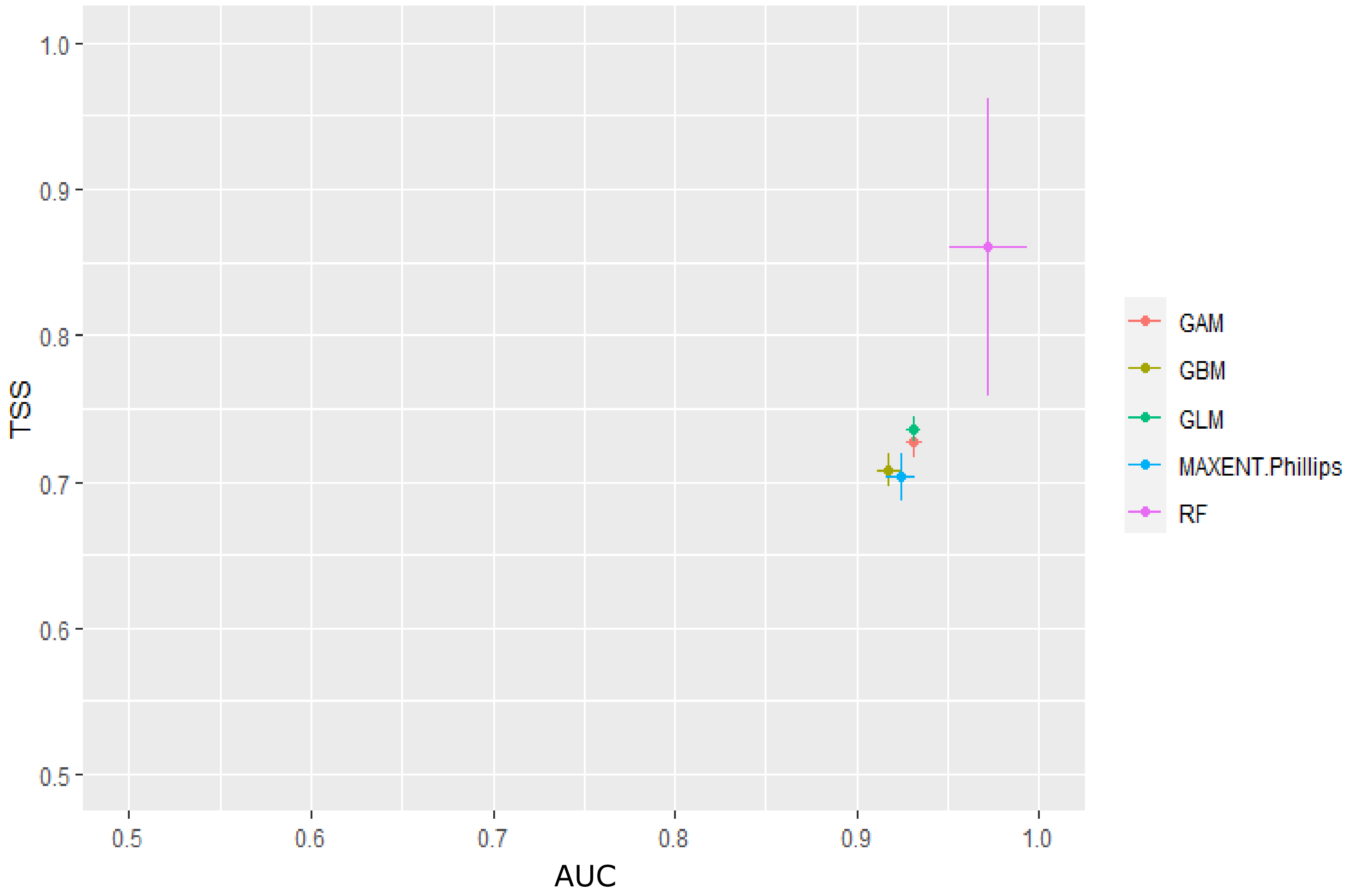




ML strategy


Predictor groups
Process
Combinations
ML algorithms
Classification strategy
Evaluation metrics
Can the information be summarized ? What is the intrinsic information carried by each group ? What group combination is the most relevant for the modelling ?
Which model gives the better results ? What process and combination are the most adatped to what model ?
•Recall: Are all of the habitats predicted ?
•Precision: How faithful are the predicted habitats ?
•F1 score: Is the balance good between the habitat prediction quantity and quality ?
•Top k accuracy: Is the real label among tho most likely ? If the recall is low, but the top k is high, how many similar habitats are co-occuring ? (given by k)
•Confusion matrix: What habitats are confused with others ? In what proportion ?
Climate
Full
Selection
PCA
RS-EBVs
Full
Selection
PCA
Soil
Full
PCA
Landcover
CLC
CLC+
Worldcover
Full density
Full dominant
PCA
Population + Distance to water
Climate
RS-EBVs
Soil
Landcover
Combination 3
Combination 2
Combination 1
PCA
Selection
Full
Full density
Full dominant
For every combination, one predictor group is put aside by turns, to assess its carried information compared to when every groups are present.
SVC
Maxent
Random Forest
XGBoost
DNN
Multi-class
Multinomial
OvR
Top 1 3 5 10
Recall
Precison
F1 score
Confusion matrix
Is multi-label favors equity in the different label predictions ? (by looking at the recall)
Multi-label
Binary relevance
Multi-task
Is the parent label information useful to classify within a given class/sub-class ?
Hierarchical (multinomial)
Level 2 (given level 1)
Level 3 (given level 2)

Differences in accuracy models from BIOMOD2



Random Forest performs with best accuracy, but takes too much time to run (> 200 hours single model). Maxent model only model possible to run at European scale

Figure Accuracy assessment for the different methods for habitat suitability modelling with same set of training data and set of predictors at 100 meter resolution. AUC = Accuracy Under the Curve. TSS = True Skill Statistics.

Validation of European habitat suitability maps using Article 17 database

Code	Description	Nr EVA plots	Overall accuracy	User's accuracy	Producer's accuracy	Commis. error	Omission error
7130	Blanket bog	822	**0.97**	0.48	0.80	0.52	0.20
9410	Acidophilous Picea forests of the montane to alpine levels (Vaccinio-Piceetea)	11042	**0.95**	0.49	0.91	0.51	0.09
6520	Mountain hay meadows	4618	**0.92**	0.40	0.66	0.60	0.34
4060	Alpine and Boreal heaths	9435	**0.92**	0.49	0.62	0.51	0.38
1510	Mediterranean salt steppes (Limonietalia)	312	**0.91**	0.09	0.69	0.91	0.31
2190	Humid dune slacks	3988	**0.91**	0.13	0.71	0.87	0.29
5120	Mountain Cytisus purgans formations	616	**0.89**	0.06	0.81	0.94	0.19
1310	Salicornia and other annuals colonizing mud and sand	17773	**0.88**	0.21	0.81	0.79	0.19
6230a	Species-rich Nardus grasslands, on silicious substrates in mountain areas (narrow sel)	1314	**0.87**	0.55	0.27	0.45	0.73
2130	Fixed coastal dunes with herbaceous vegetation (grey dunes)	8927	**0.85**	0.16	0.83	0.84	0.17
4010	Northern Atlantic wet heaths with Erica tetralix	2081	**0.83**	0.20	0.93	0.80	0.07
9110	Luzulo-Fagetum beech forests	2906	**0.79**	0.39	0.72	0.61	0.28
6230b	Species-rich Nardus grasslands, on silicious substrates in mountain areas (broad sel)	10828	**0.76**	0.33	0.78	0.67	0.22
9180	Tilio-Acerion forests of slopes, screes and ravines	6541	**0.68**	0.31	0.79	0.69	0.21
3230	Alpine rivers and their ligneous vegetation with Myricaria germanica	554	**0.67**	0.02	0.98	0.98	0.02
3240	Alpine rivers and their ligneous vegetation with Salix elaeagnos	2343	**0.64**	0.08	0.99	0.92	0.01
6410	Molinia meadows on calcareous, peaty or clayey-silt-laden soils (Molinion caeruleae)	8220	**0.56**	0.29	0.80	0.71	0.20
7110	Active raised bogs	3640	**0.54**	0.18	0.96	0.82	0.04
6210	Semi-natural dry grasslands and scrubland facies on calcareous substrates	646	**0.48**	0.29	0.91	0.71	0.09
8210	Calcareous rocky slopes with chasmophytic vegetation	2018	**0.43**	0.19	0.86	0.81	0.14
8160	Medio-European calcareous scree of hill and montane levels	827	**0.42**	0.03	1.00	0.97	0.00
8220	Siliceous rocky slopes with chasmophytic vegetation	526	**0.38**	0.18	0.57	0.82	0.43
5130	Juniperus communis formations on heaths or calcareous grasslands	879	**0.35**	0.07	0.95	0.93	0.05

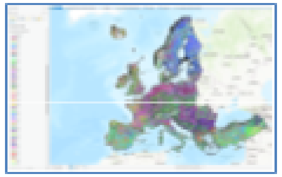
Integration European habitat suitability maps towards wall-to-wall mapping

…

Overlay all habitat suitability maps for a specific habitat group (in this case forests) and extract habitat types with highest suitability (python scripts)

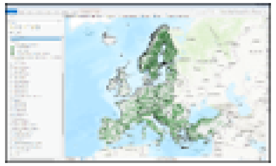
Forest types with the highest suitability per pixel (Max1_Foresttypes)

Forest types with the second best suitability score per pixel (Max2_Foresttypes)

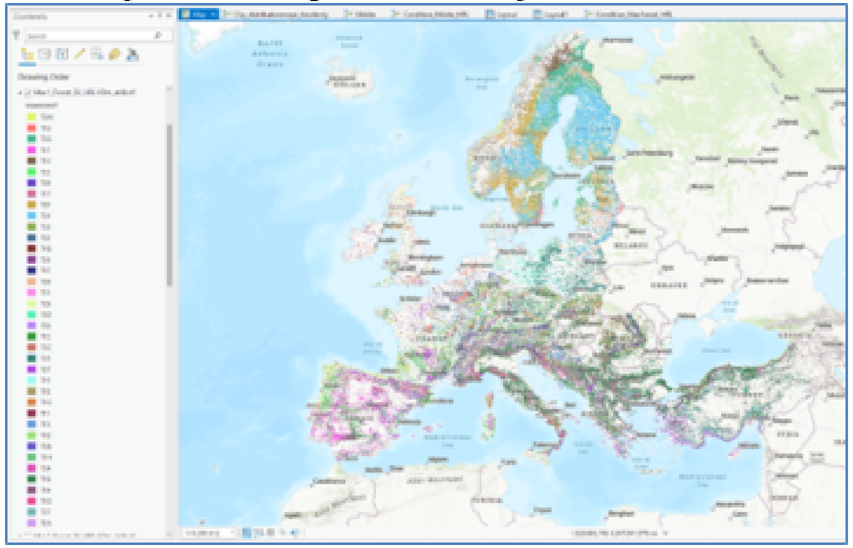


Forest types with the third best suitability score per pixel (Max3_Foresttypes)

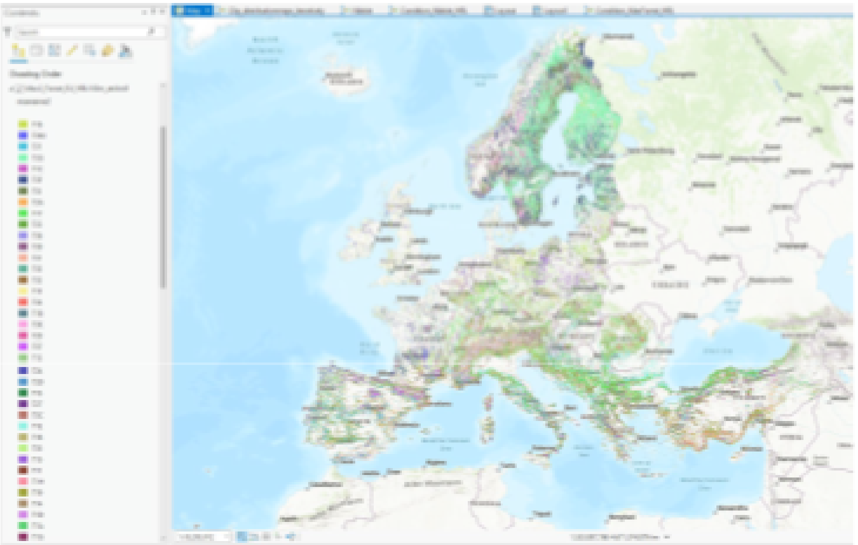
Copernicus HR layer Forest (EEA)



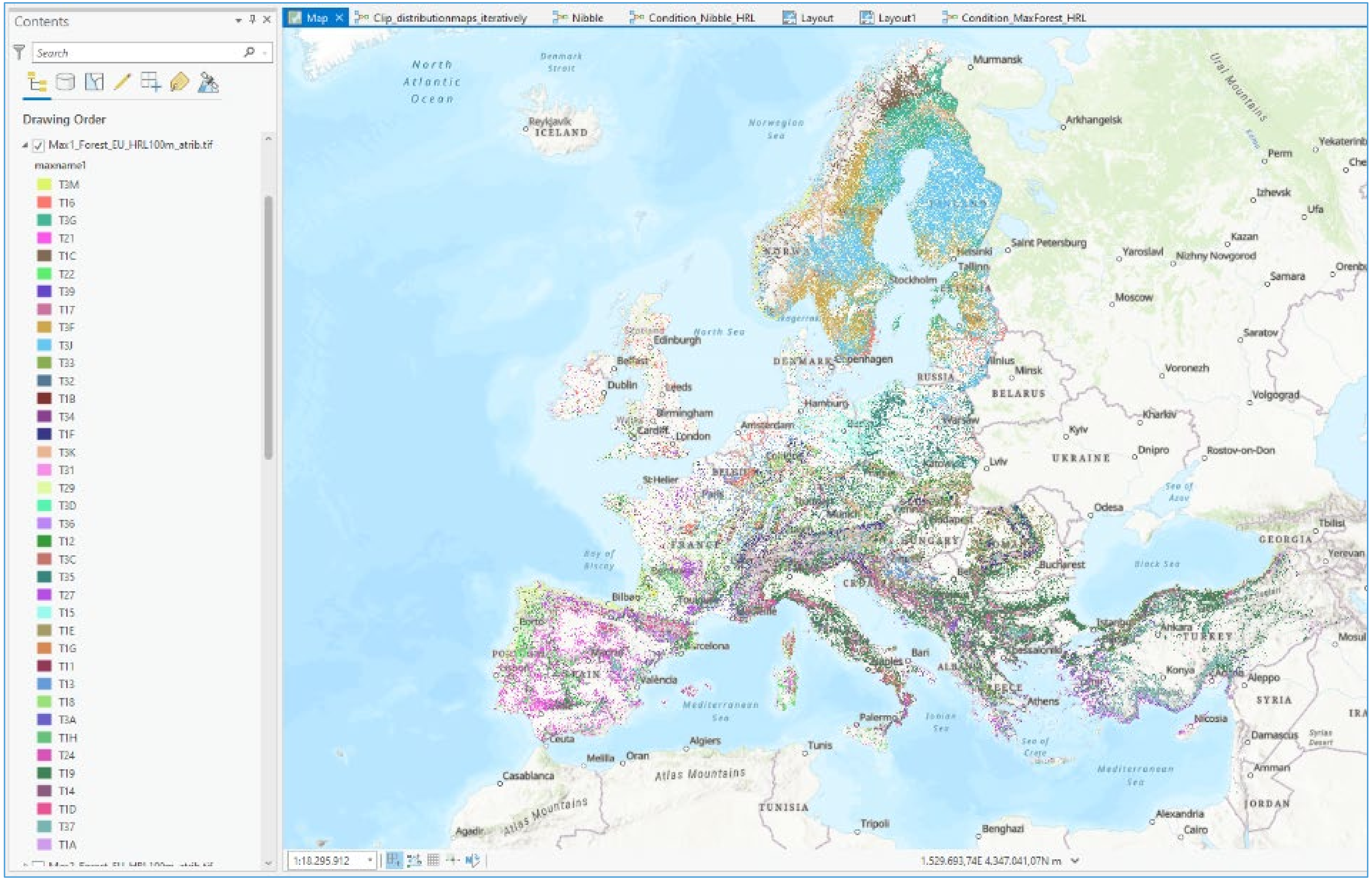
Clip the integrated EUNIS Forest habitat map (level 3) with the highest suitability with the Copernicus HR layer



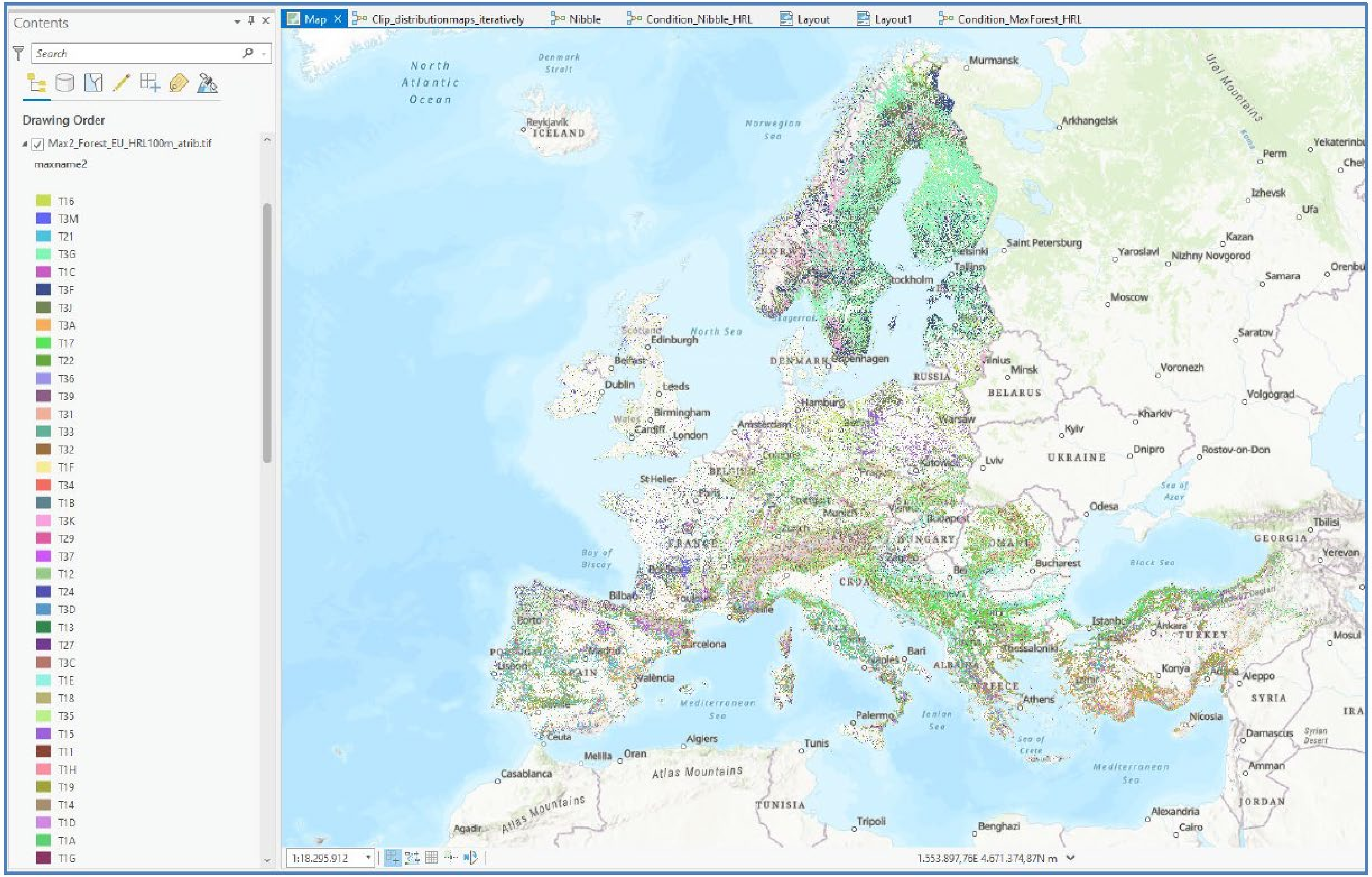
Clipped forest types with the second best suitability score per pixel (Max2_Foresttypes)



EUNIS Forest habitat map (level 3) with the highest suitability score for forest habitats clipped with the Copernicus HR layer Forest.

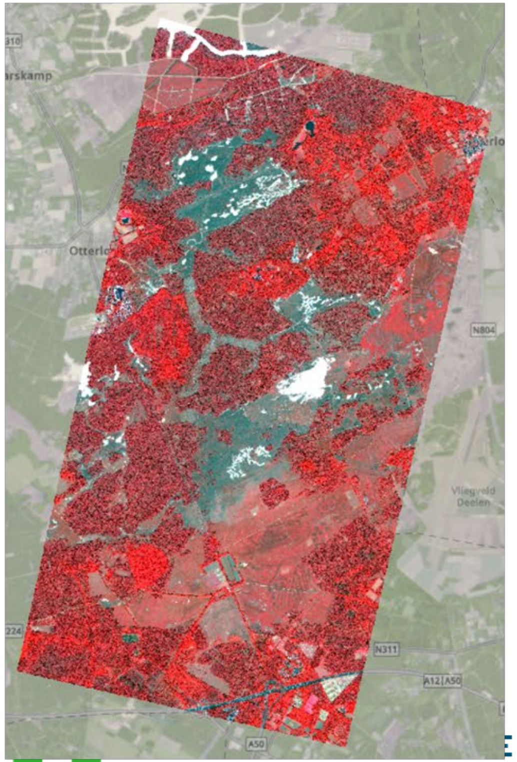


Integration: EUNIS Forest habitat map (level 3) with the second best suitability score for forest habitats clipped with the Copernicus HR layer Forest.

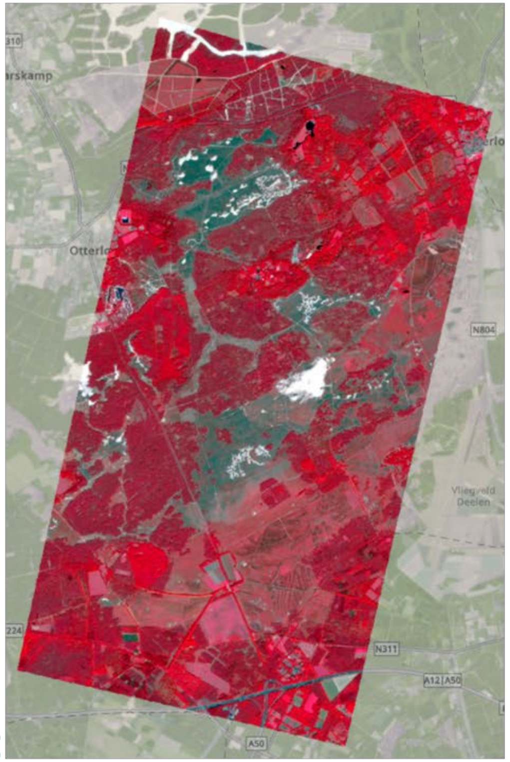


Method 2 Regional habitat mapping using deep learning techniques

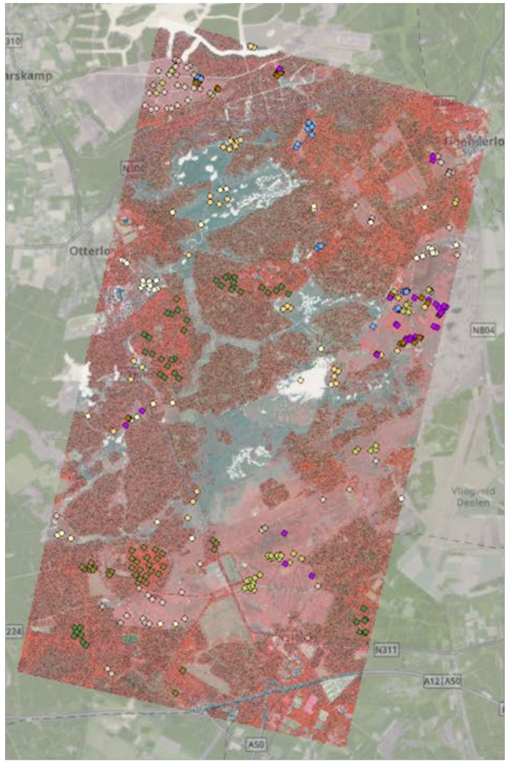
- WENR & VITO are working on exploitation deeplearning models for habitat mapping at regional and national scale. For example in National Park Veluwe, the Netherlands, using HR-VPP and Sentinel-2 at 10 meter resolution (next to Superview)



Superview 12-08-2020, False colour



Sentinel 31-07-2020, False colour



Selected LVD points in Hoge Veluwe test area

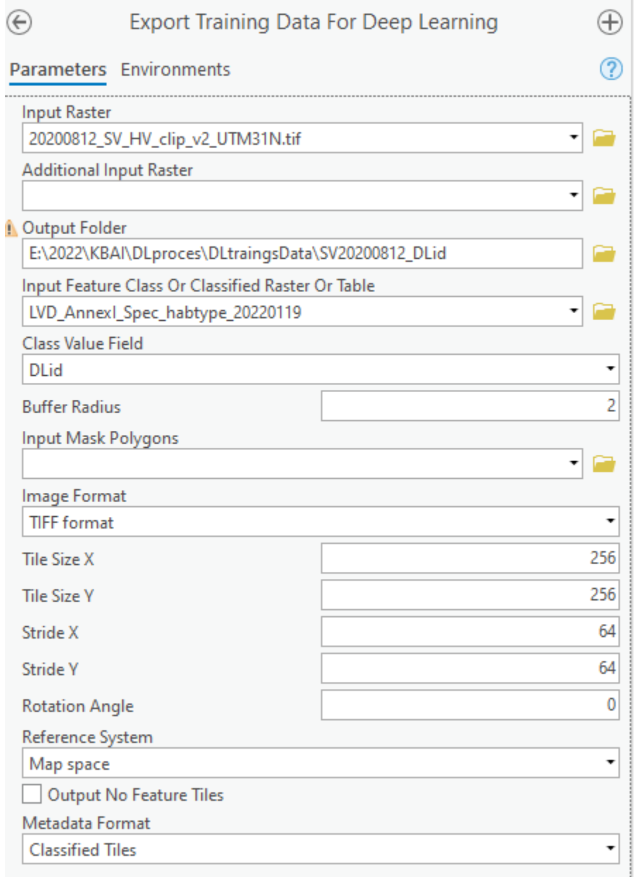
	SPEC_HABTY,DLid	
24	23101,1	Dry sand heaths (light) - sand
16	23102,2	Dry sand heaths (dark) - vegetated
4	23301,3	Inland dunes (light)
34	23302,4	Inland dunes (dark)
31	31601,5	Lakes and ponds
40	40101,6	Wet heaths
44	40301,7	European dry heaths (light) - Pijpenstrootje
19	40302,8	European dry heaths (dark) - heide
40	62301,9	Species-rich Nardus substrates
39	71501,10	1 Depressions on peat substrates
37	91201,11	1 Birch forests
45	91901,12	1 Oak woods

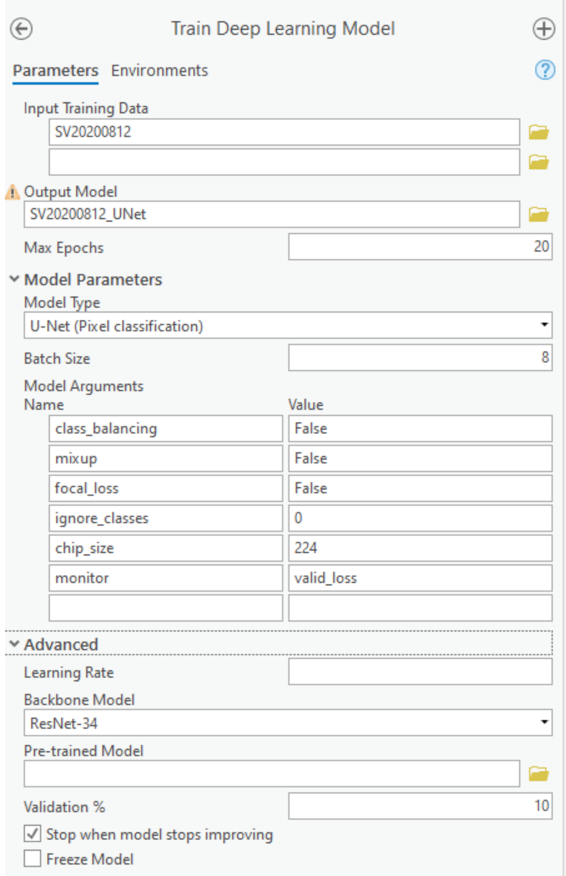
Method 2a Deep Learning (U-NET in ArcGIS PRO)

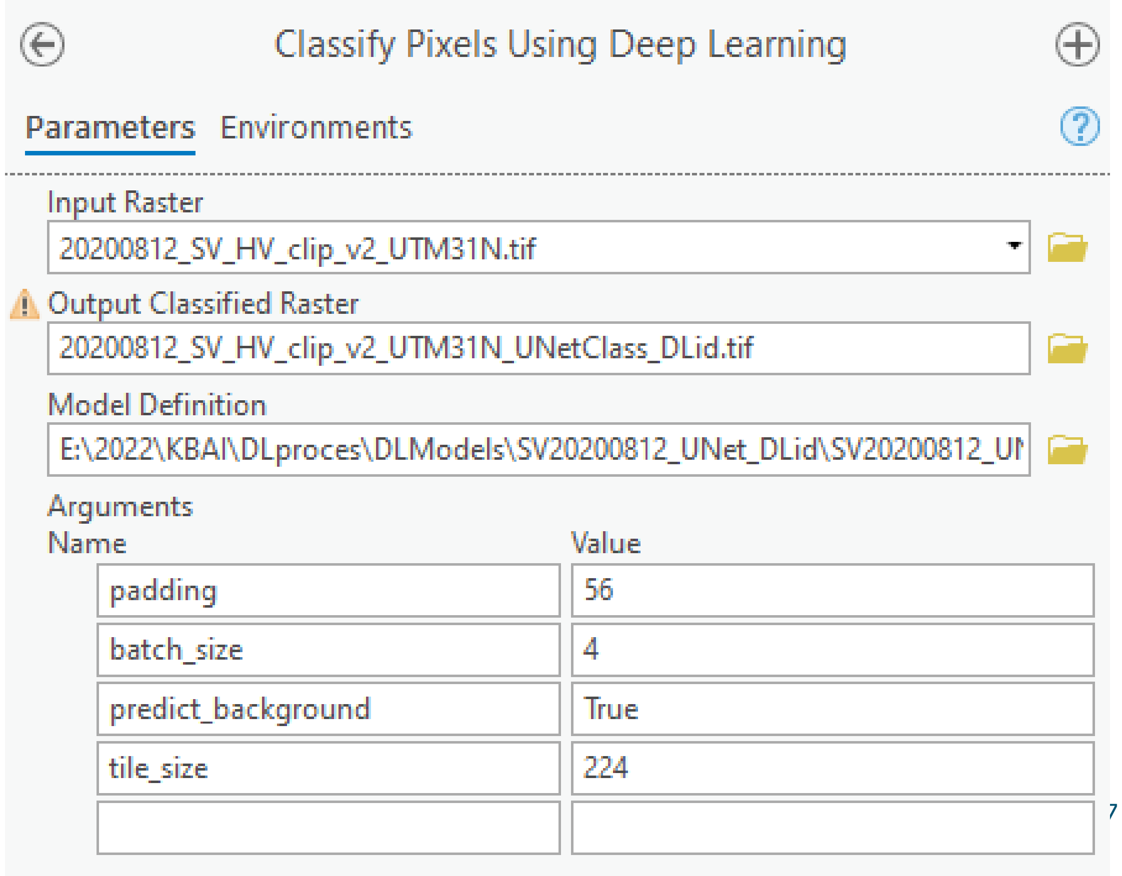
Step 1
Prepare Training Data

Step 2
Train a Model

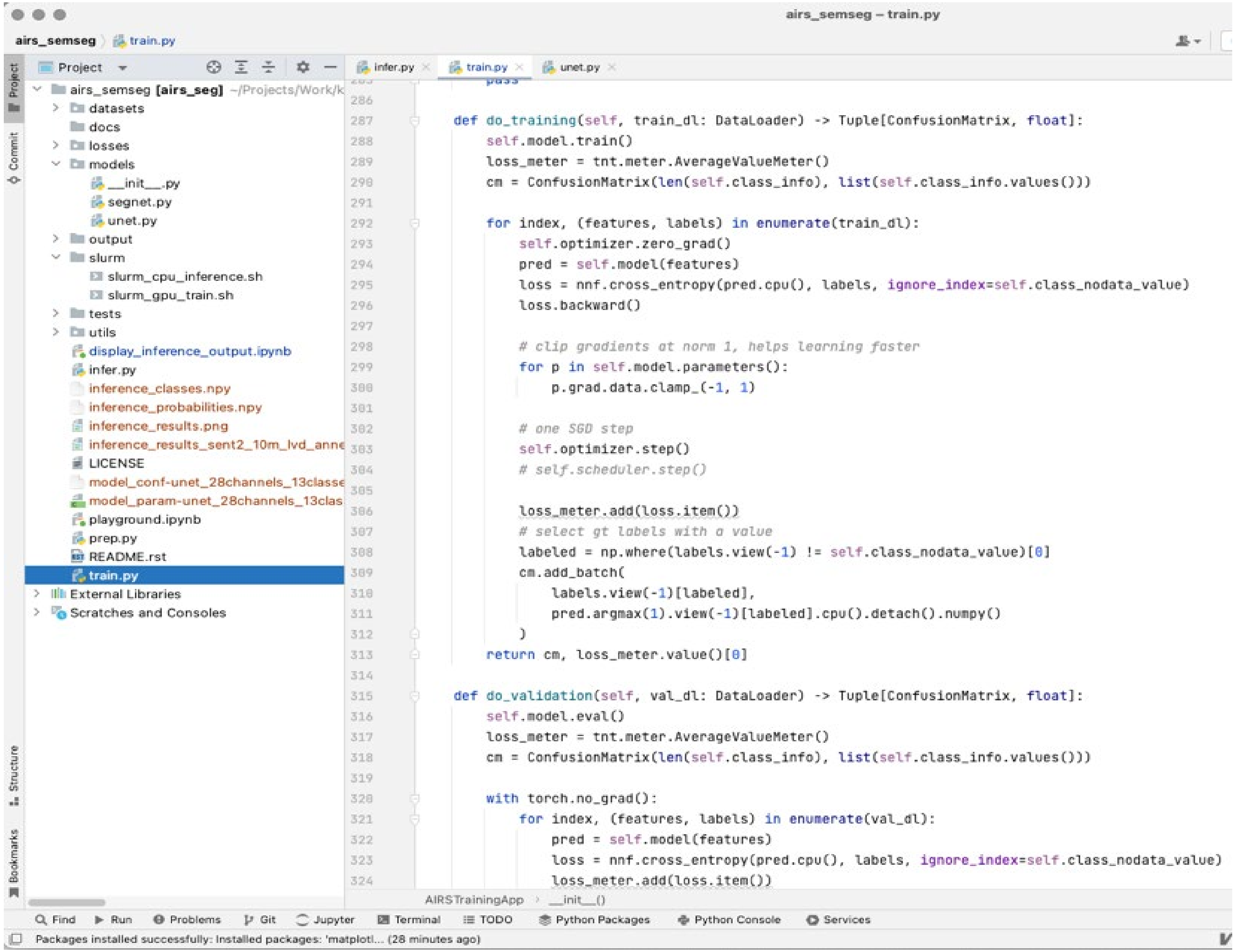
Step 3
Use the Model







Method 2b Deep Learning in Pytorch with U-NET

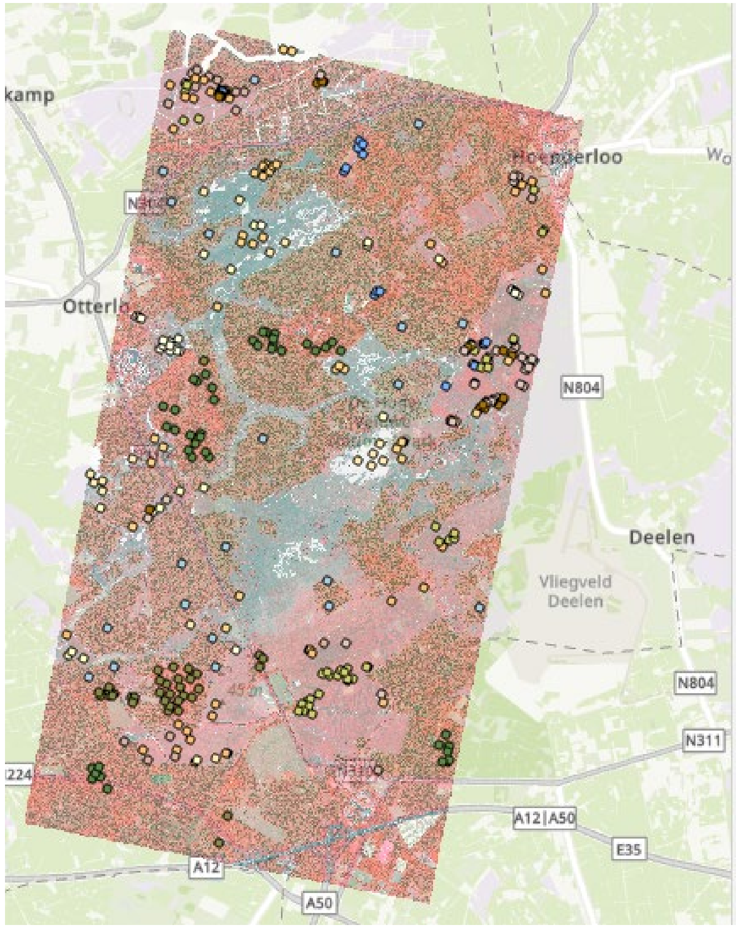


PyTorch and TorchVision based

- Steps:
  - prep.py: Create HDF5 tensor inputs
  - train.py: Train DL model
  - infer.py: Apply DL model
- Training needs GPUs, can be done on Anunna HPC or Cloud, e.g. Azure.
- Choice of DL models, starting with regular U-Net.
- Various Remote Sensing data:
  - Sentinel 2 (10-20m)
  - SuperView (50 cm)

Deep Learning proces in ArcGIS PRO

Selected training points for Deep Learning process from Landelijke Vegetatie Database (LVD)



Classes | Scales

DLid — 13 symbol classes

Symbol	Value	Label	Count
○	1	1 Dry sand heaths (light) - sand	19
○	2	2 Dry sand heaths (dark) - vegetated	16
○	3	3 Inland dunes (light)	16
○	4	4 Inland dunes (dark)	33
○	5	5 Lakes and ponds	28
○	6	6 Wet heaths	39
○	7	7 European dry heaths (light) - Pijpenstrootje	35
○	8	8 European dry heaths (dark) - heide	19
○	9	9 Species-rich Nardus substrates	37
●	10	10 Depressions on peat substrates	33
●	11	11 Birch forests	37
●	12	12 Oak woods	44
○	13	13 Coniferous forest	21

Habitat types Dry sands heaths (2310), Inland dunes (2330) and European dry heaths (4030) were divided into two subclasses each because for these three habitat types both light and dark appearances in the satellite image can be seen.

All training points were checked on their class and geometric validity and edited if necessary. Additional points for Inland dunes (light) were digitized because there were only four points available from the LVD.

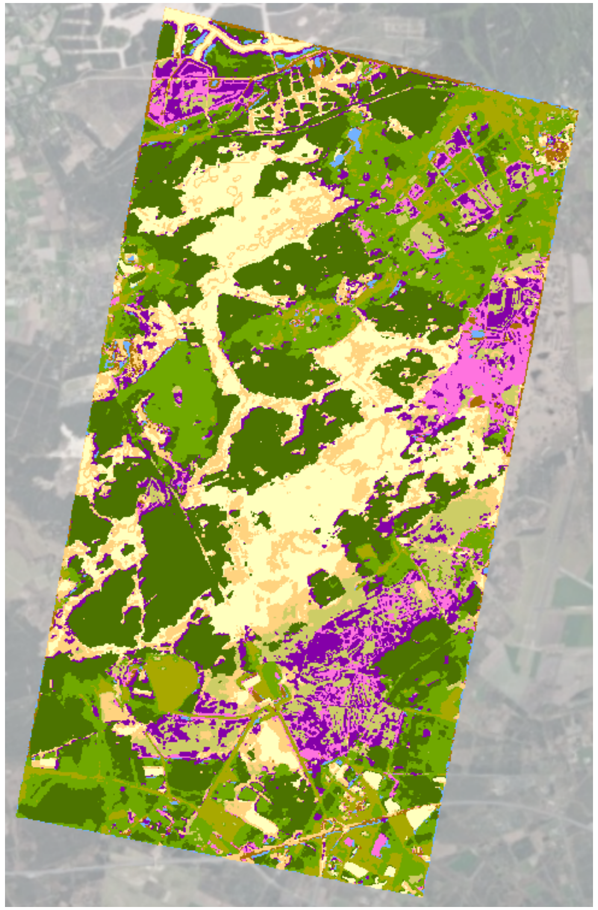
WAGENINGEN UNIVERSITY & RESEARCH

U-NET deep learning technique

Groundtruth habitat map

Habitat map

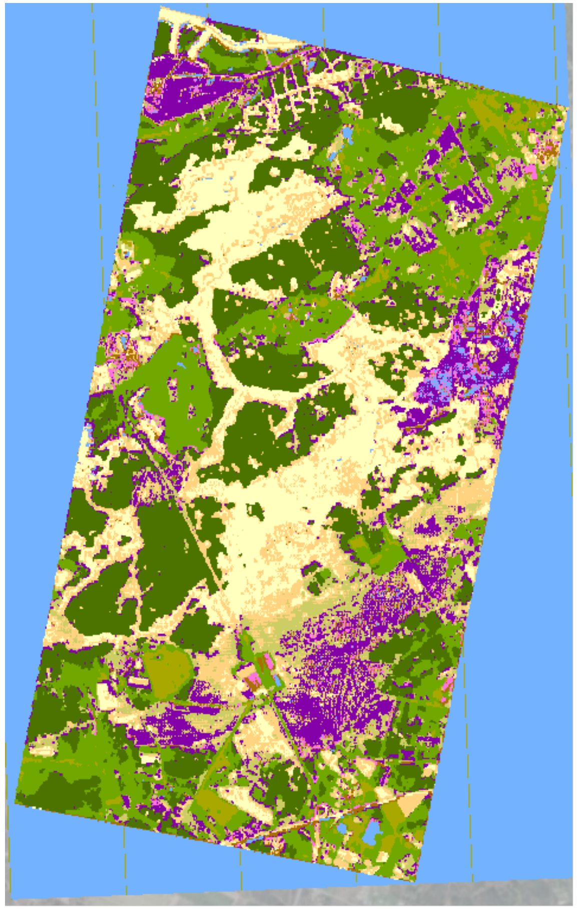
Result U-Net in ArcGIS Pro



Sentinel 2020 stack 7 images

07-02 07-05 14-09
23-03 26-06
15-04 31-07

Result U-Net in Pytorch



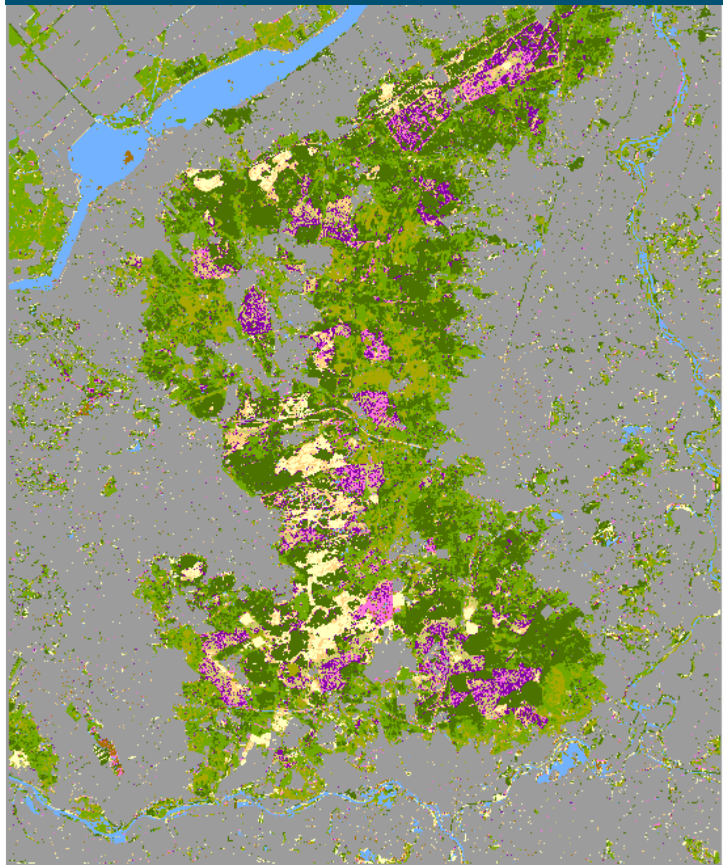
Sentinel 2020 stack 7 images
Classification in Pytorch

WAGENINGEN
UNIVERSITY & RESEARCH

KB-DDHT-2 AI for Remote Sensing

Upscaling trained DL model subset Hoge Veluwe to classification entire Veluwe using

Sentinel 2020 - 7 images

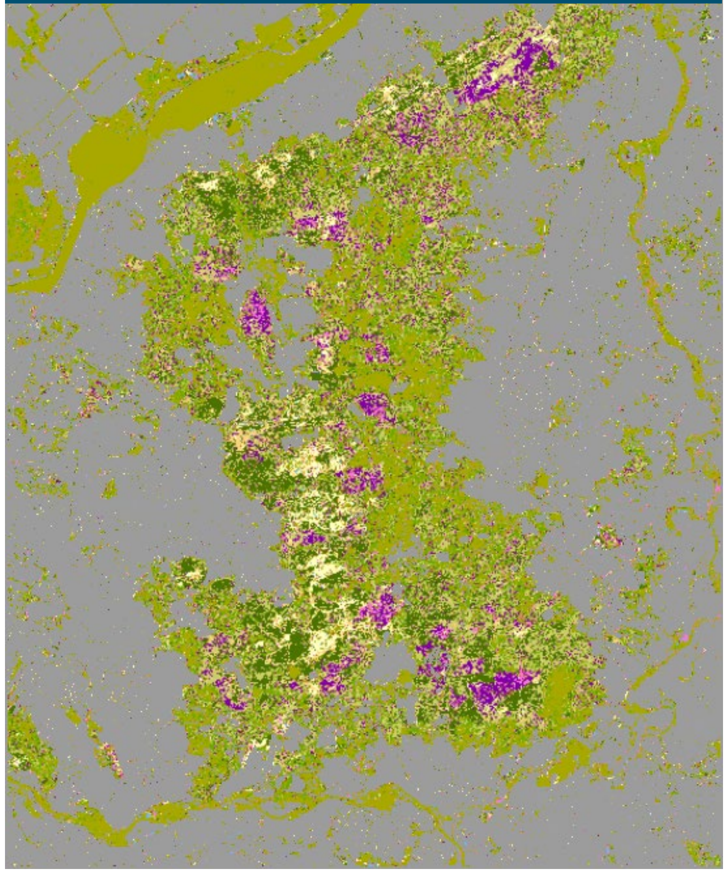


Deeplearning
with U-NET

Habitat type

- 2310 Dry sand heaths
- 2330 Inland dunes
- 3160 Lakes and ponds
- 4010 Wet heaths
- 4030 European dry heaths
- 6230 Species-rich Nardus substrates
- 7150 Depressions on peat substrates
- 9120 Beech forests
- 9190 Oak woods
- Coniferous forest

HR-VPP (High Resolution–Vegetation Phenology Product) – Sentinel 2



WAGENINGEN UNIVERSITY & RESEARCH

Sentinel 2020 stack 7 images: 07-02, 23-03, 15-04, 07-05, 26-06, 31-07, 14-09

Validation

		classification result												
LDV training points	**HABITATTYP**	**2310**	**2330**	**3160**	**4010**	**4030**	**6230**	**7150**	**9120**	**9190**	**s forest**	**total**	**producer accuracy**	
dry sand and heaths	2310	23	5		1	1	5					35	66%	
inland dunes	2330	3	42		1		2	1				49	86%	
lakes and ponds	3160			26	1			1				28	93%	
wet heaths	4010	1			27	7		4				39	69%	
european dry heaths	4030	1			4	40	9					54	74%	
species-rich nardus substrates	6230				3	5	28			1		37	76%	
depression on peat substrates	7150				10	2		21				33	64%	
beech forest	9120								34	3		37	92%	
oak woods	9190						1		1	42		44	95%	
coniferous forest	13										21	21	100%	
	Grand total	**28**	**47**	**26**	**47**	**55**	**45**	**27**	**35**	**46**	**21**	**377**		
	user accuracy	82%	89%	100%	57%	73%	62%	78%	97%	91%	100%		**81%**	**overall accuracy**

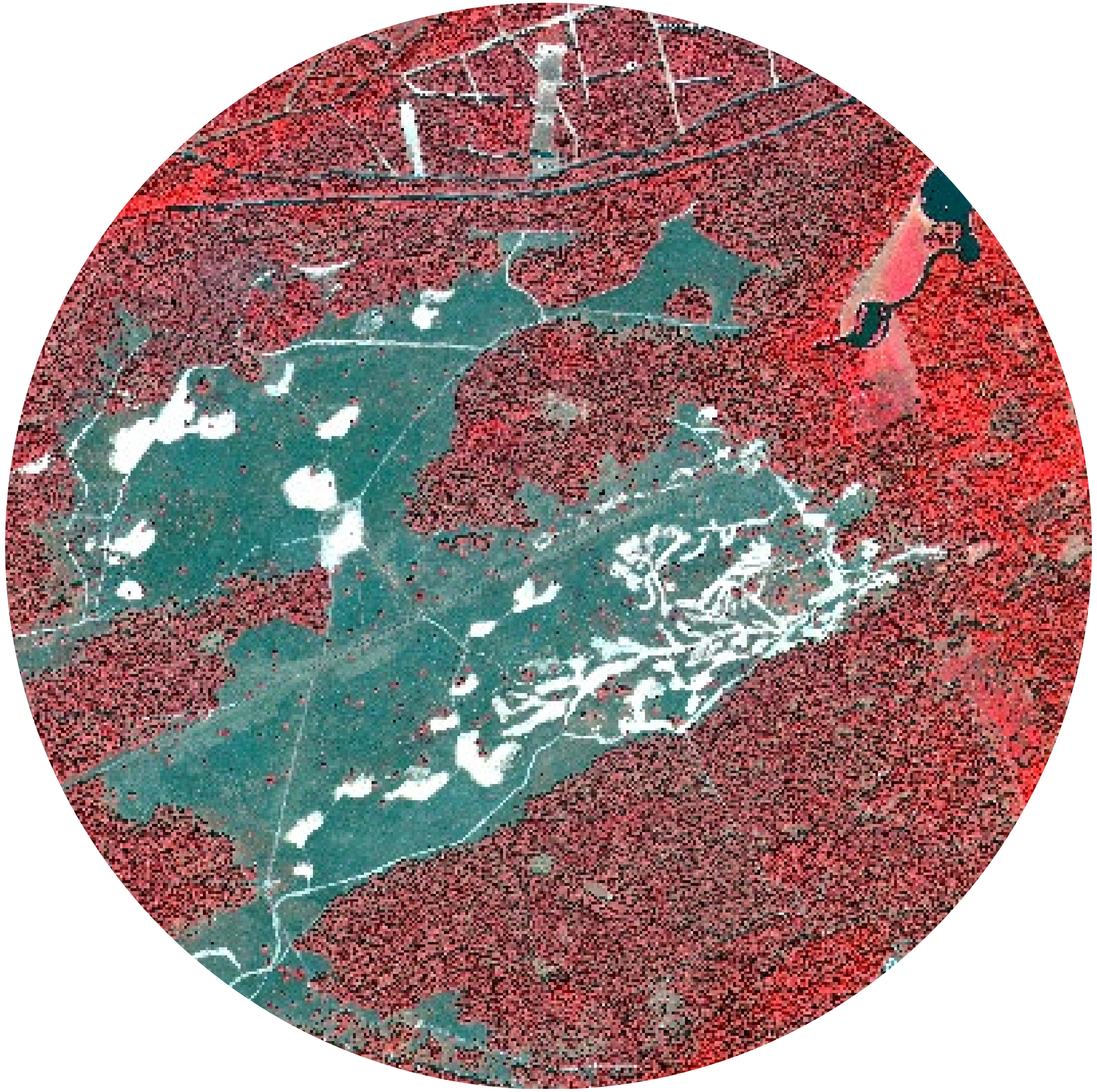
Conclusions & discussion

- We can model the suitability for 203 EUNIS habitats with Maxent model across Europe with reasonable results
- Validation of European habitat maps show in general high overall accuracies, but is mainly due to a large amount of true negatives.
- To improve the user accuracy it is necessary to refine the European habitat suitability maps with accurate land cover maps.
- Integration of the individual European habitat suitability maps for wall-to-wall mapping could also be improved by using a multiclass ML approach.
- With deep learning techniques on satellite imagery we are able to map European habitats at regional scale using both our own Pytorch scripts as well with ArcGIS Pro. But there is still much room for improvements.
- But upscaling with DL techniques requires much of data infrastructure and CPU en GPU capacity.
- Habitat mapping with deep learning techniques on remote sensing imagery is the future and needs to be exploited further

Results & conclusions

- But upscaling with DL techniques requires much of data infrastructure and CPU en GPU capacity.
- Habitat mapping with deep learning techniques on remote sensing imagery is the future and needs to be exploited further
- Amount and quality of training data is crucial. Much time goes in the enhancement of training data.
- Selection of vegetation plot data (from e.g. EVA) for training AI/ML is more difficult than often thought – due to inaccuracies in locations
- Enhancement of the training data is a crucial step that needs much attention !!

Thank you for your attention



Contact person

sander.mucher@wur.nl

WAGENINGEN
UNIVERSITY & RESEARCH