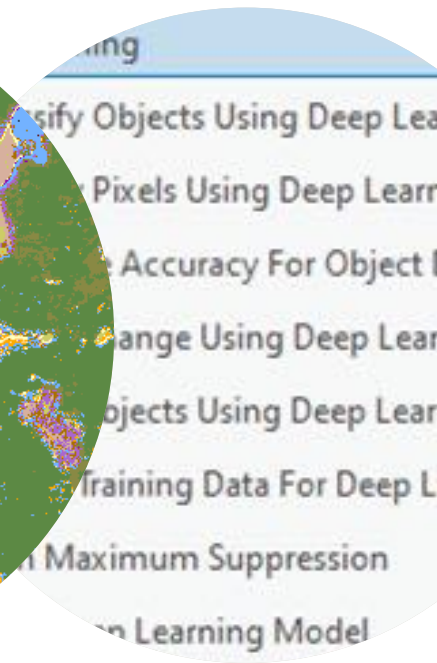
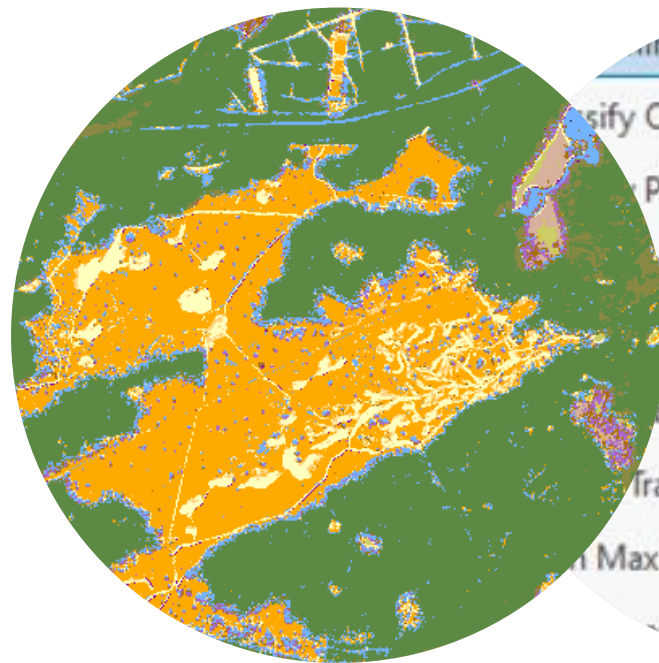


Modelling and Mapping Habitats at European and Regional Scale using AI/ML techniques

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

Sander Mucher¹; Stephan Hennekens¹; Bruno Smets²; Sara Simoussi³; Henk Kramer¹; Rob Knapen¹; Marcel Buchhorn²; Wilfried Thuiller³; Kristof Vantricht²; Stan Los¹

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



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

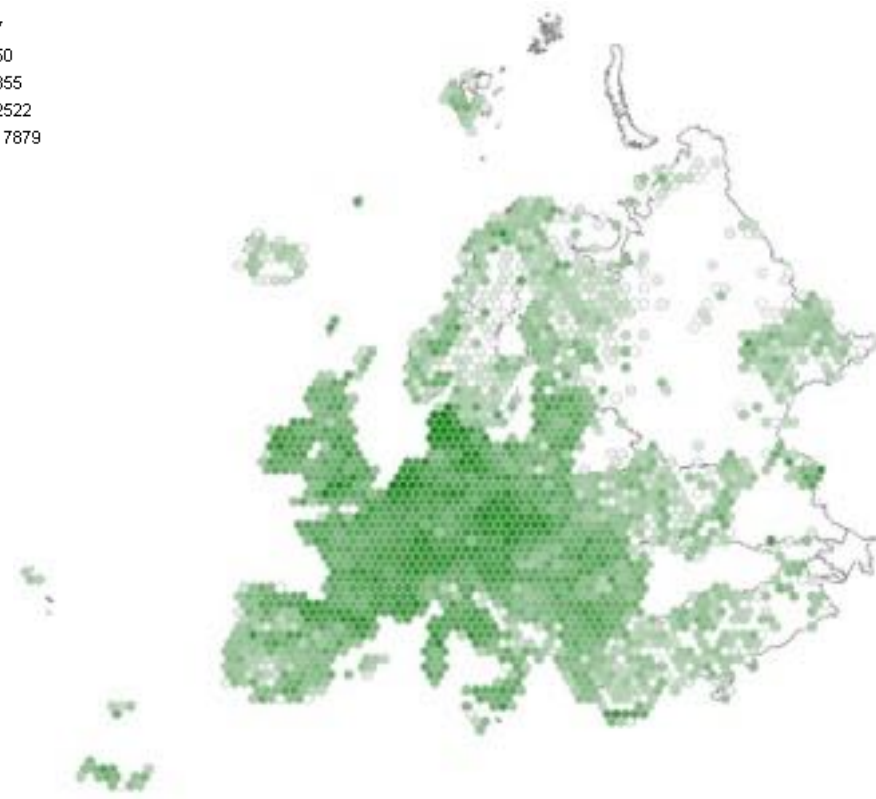
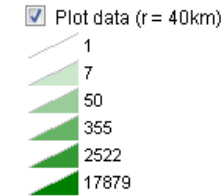
Two approaches:

- 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)



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 ($PP I \times day^{-1}$)	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 - coarse 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

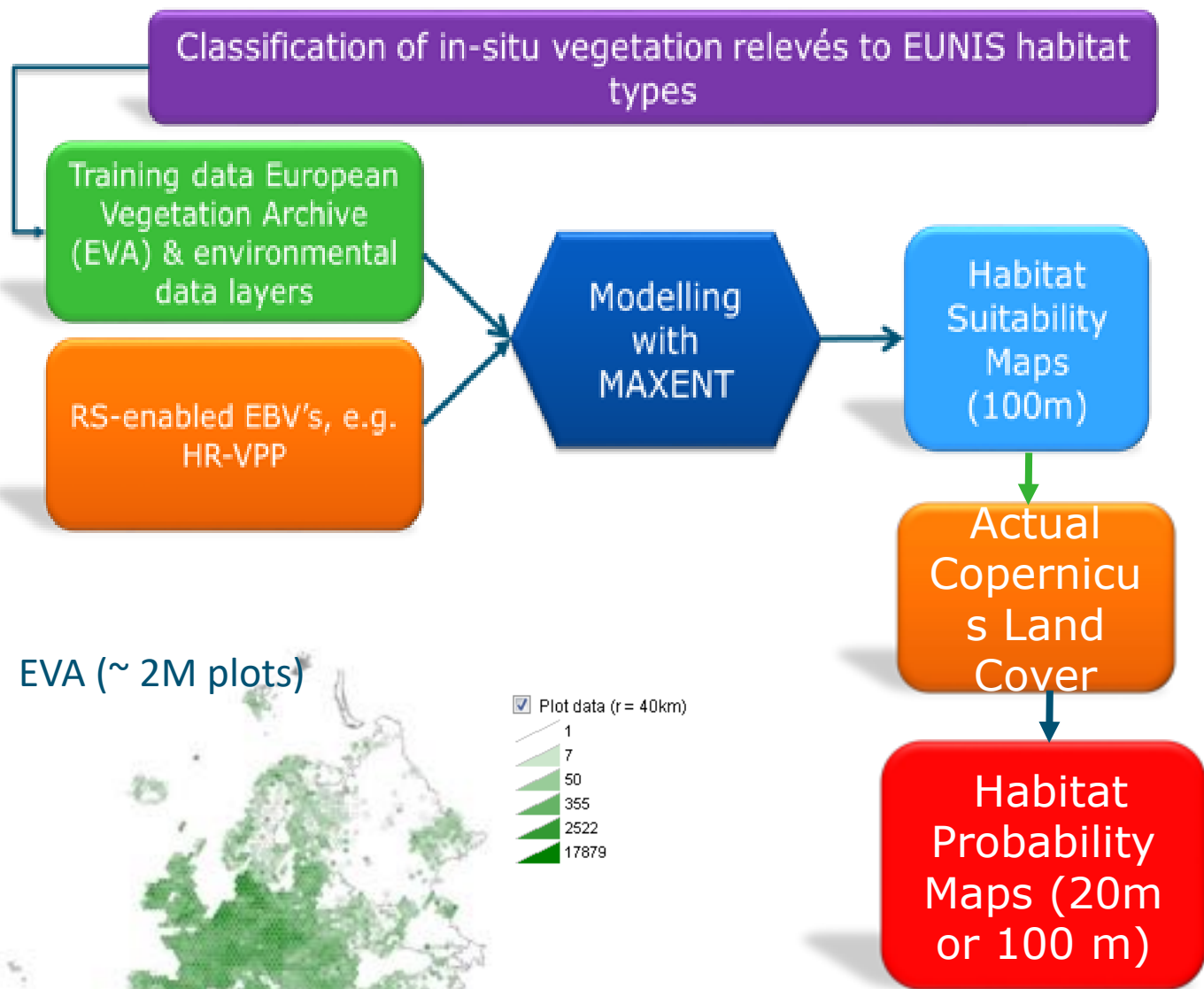


Statistics from Maxent modelling

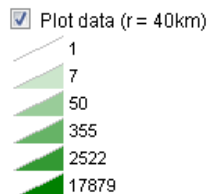
AUC training (0-1)	0.8679
AUC test (0-1)	0.8384
10 percentile training presence threshold (0-1)	0.2797
Contribution variables to the Maxent model (%)	
Climate - Snow covered days (scd)	30.63
Soil - pH	29.7015
Corine Land Cover 2018	10.8223
Climate - Accumulated precipitation amount on growing s	8.8832
EU DEM	7.2818
Climate - Growing degree days heat sum above 5°C (gdc)	4.3854
Soil - clay fraction	3.2767
Soil - coarse 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 TREI	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 st)	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



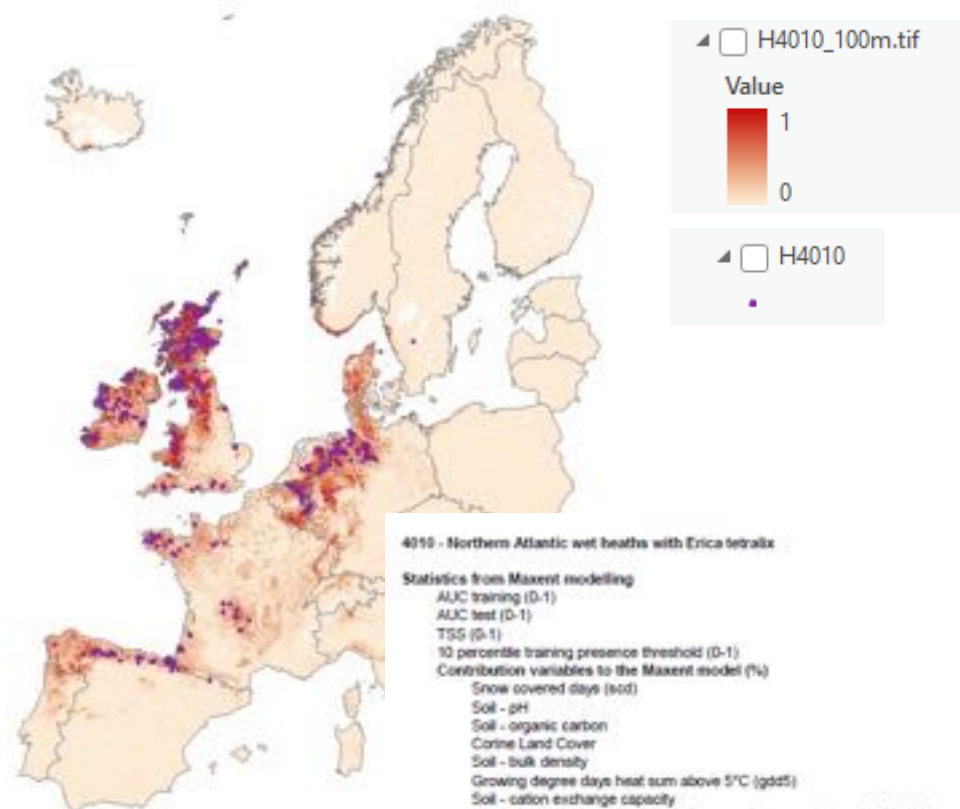
Flowchart European habitat modelling



EVA (~ 2M plots)



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 (D-1)	0.9018
AUC test (D-1)	0.9006
TSS (D-1)	0.8455
10 percentile training presence threshold (D-1)	0.3927
Contribution variables to the Maxent model (%)	
Snow covered days (sod)	39.0162
Soil - pH	27.9856
Soil - organic carbon	15.1298
Cornine 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 on growing season days TREEIM (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 (PP1 × day ⁻¹)	0.0682
Soil - course fractions	0.0491
Mean temperature of the growing season TREEIM (gdt)	0.0258
Distance to inland water	0.0182
HR-VPP - Season amplitude given by MAX-MIN/	0.017
HR-VPP - PPI at the day of maximum-of-season	0.013
Inundation - occurrence	

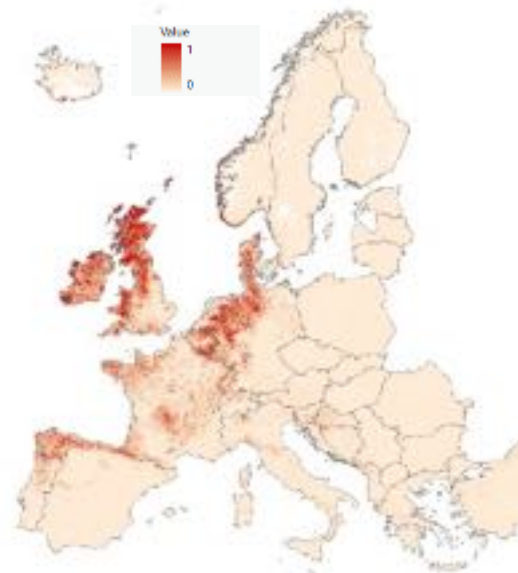
Example European habitat modelling: S41 Wet heath

Distribution data

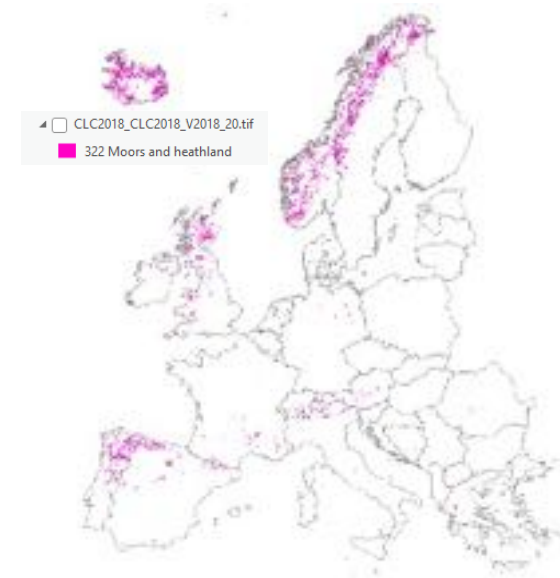
from European Vegetation Archive (EVA)



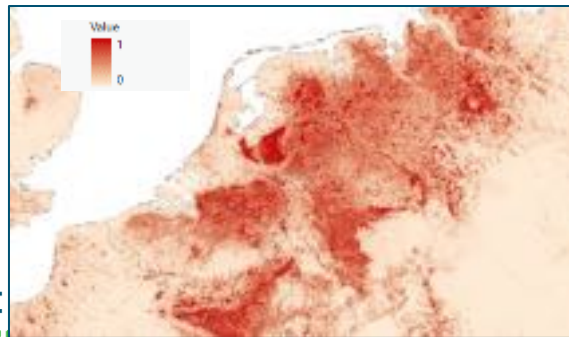
Habitat suitability map



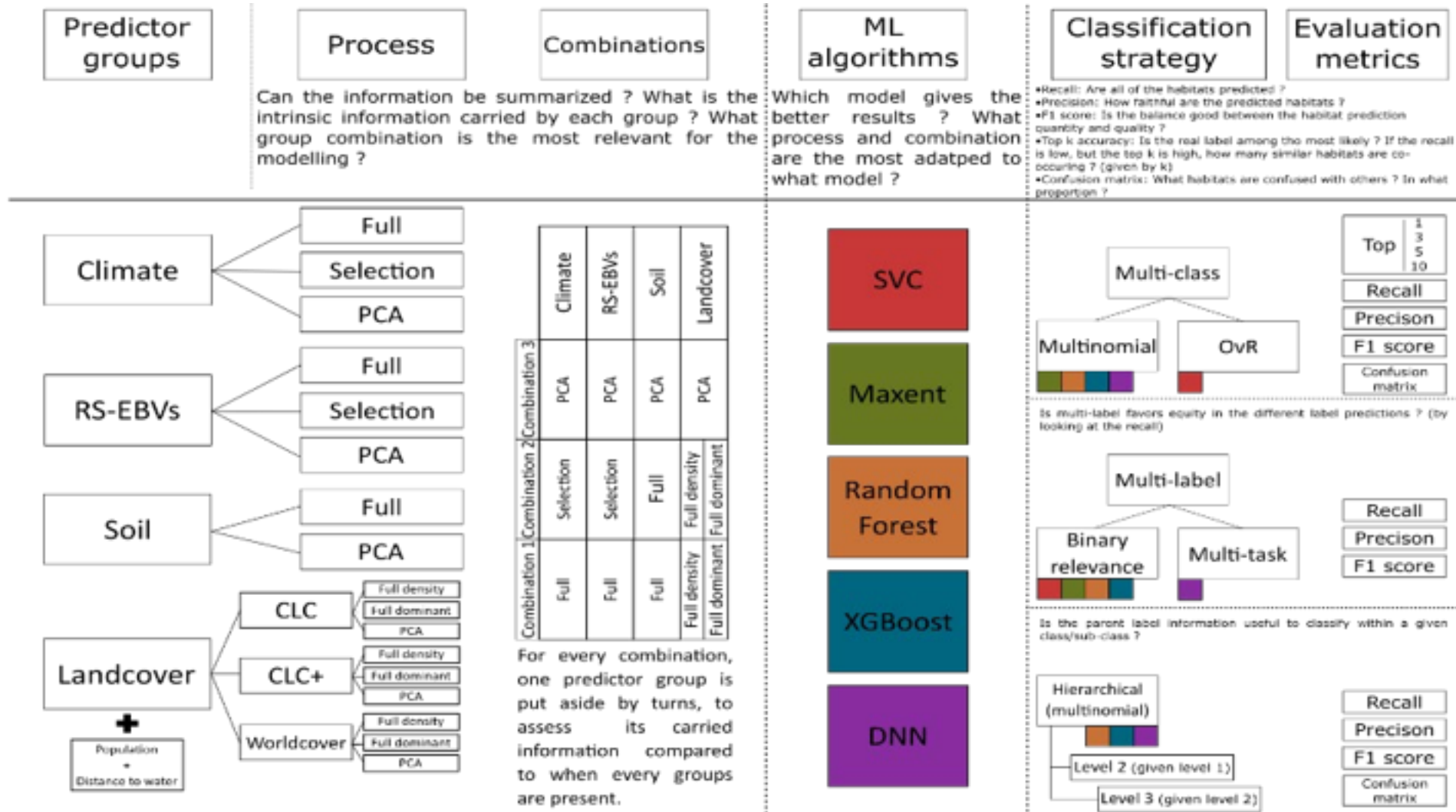
Land cover



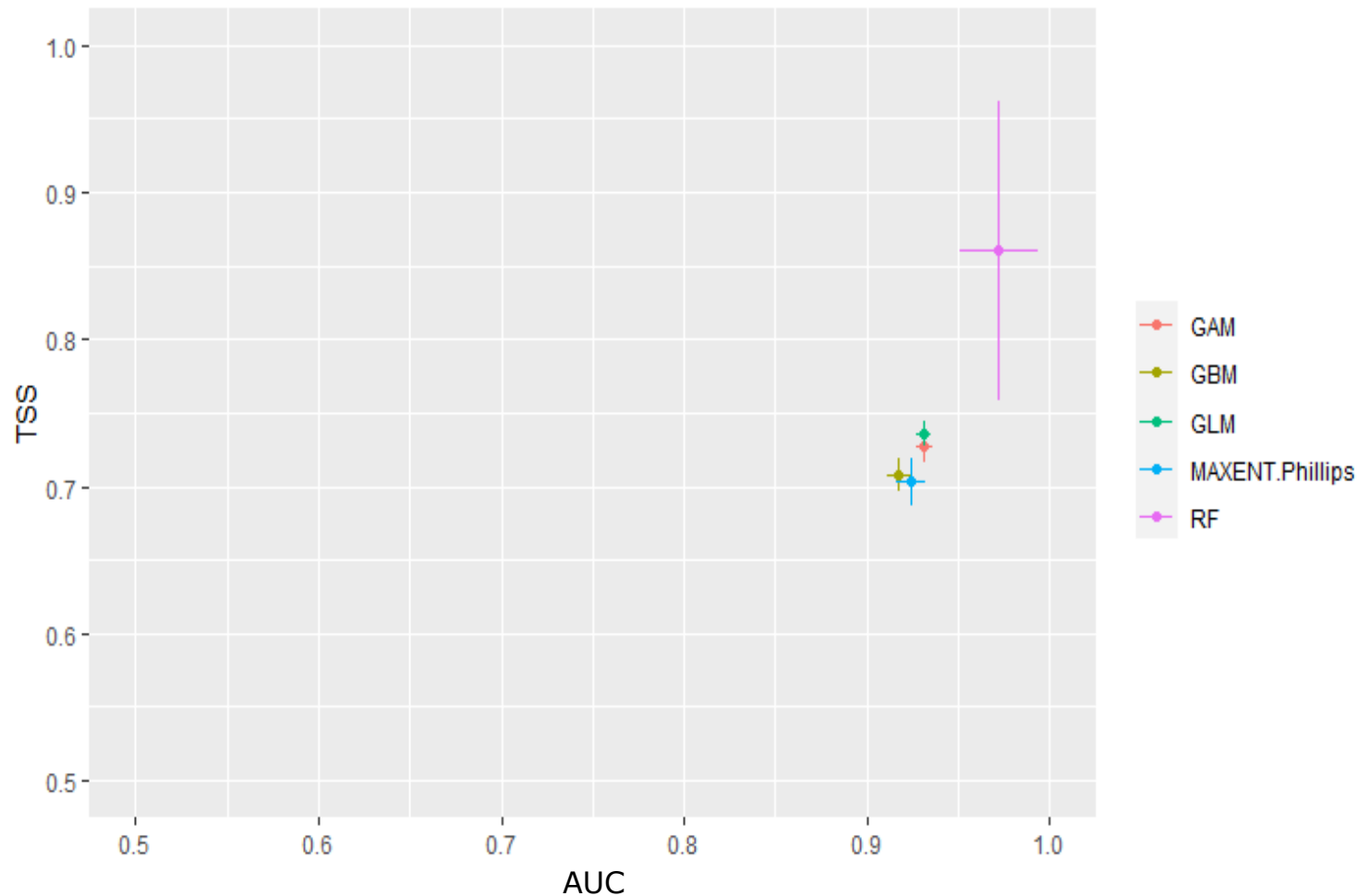
Habitat probability map



ML strategy



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	0.43.2
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)



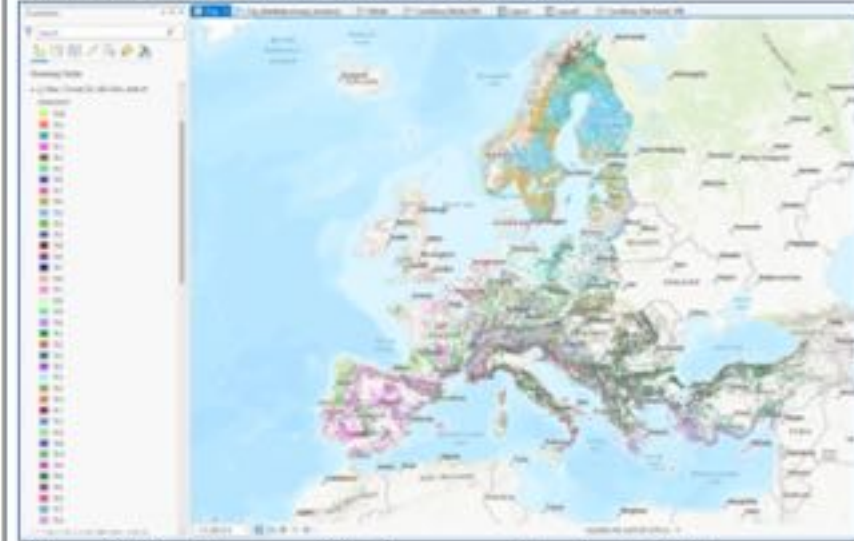
Forest types with the third best suitability score per pixel (Max3_Foresttypes)



Copernicus HR layer Forest (EEA)



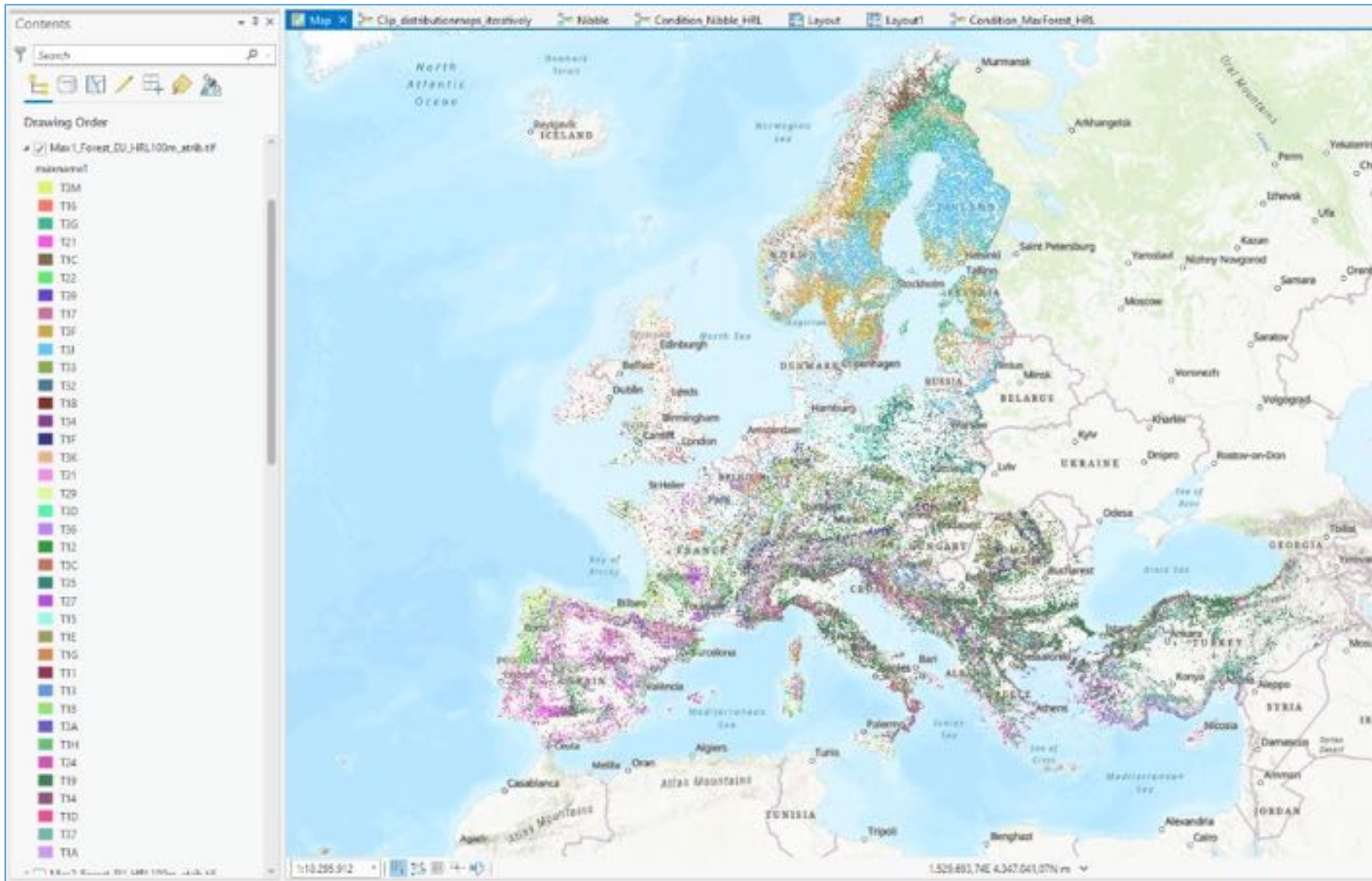
Clip the integrated EUNIS Forest habitat map (level 3) with the highest suitability with the Copernicus HR layer



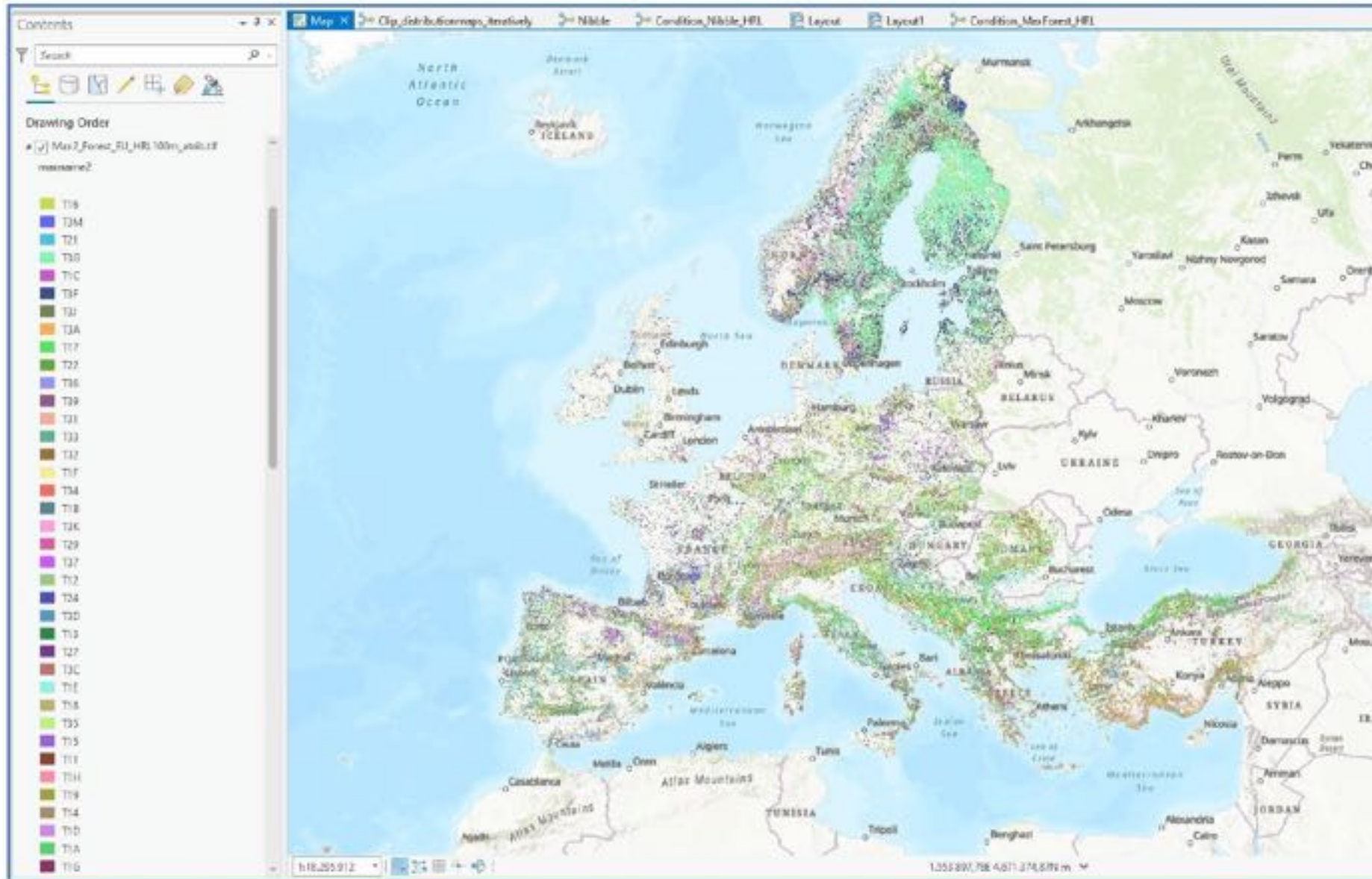
Clipped forest types with the second best suitability score per pixel (Max2_Foresttypes)



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

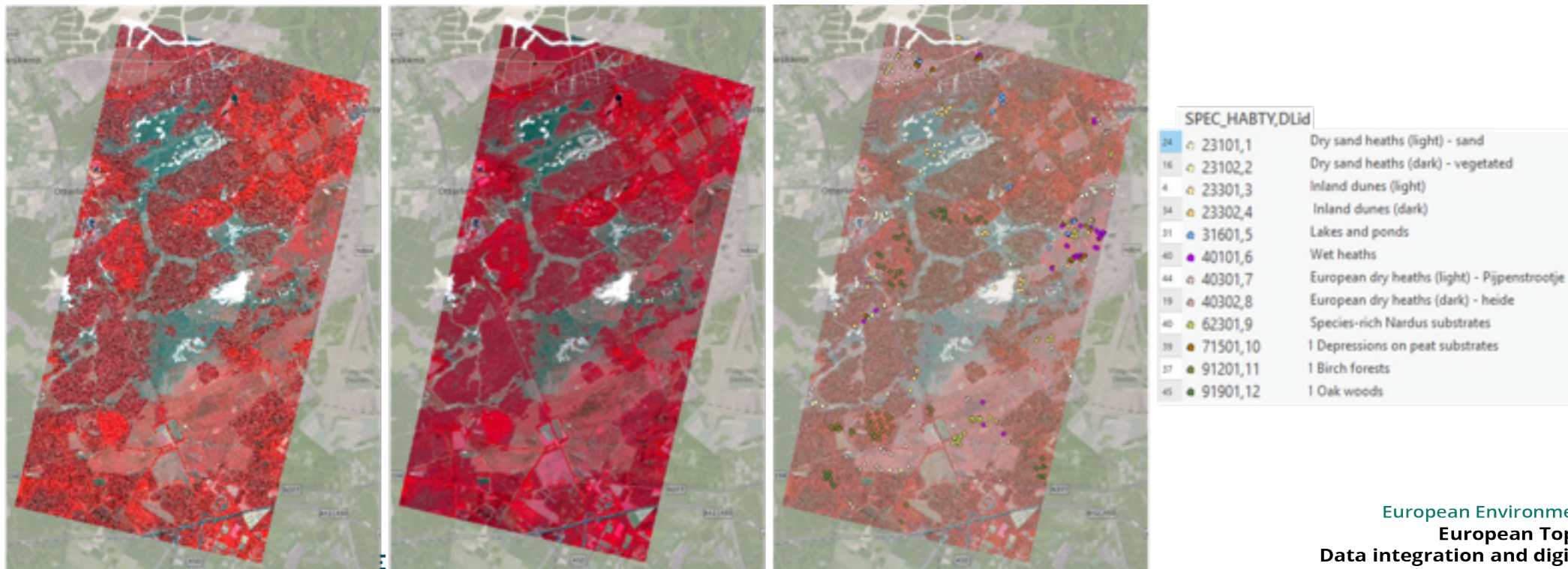


Integration: EUNIS Forest habitat map (level 3) with the second best suitability score for forest habitats clipped with the Copernicus HR layer Forest.



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



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

Step 1 Prepare Training Data

Step 2 Train a Model

Step 3 Use the Model

Export Training Data For Deep Learning

Parameters Environments

Input Raster
20200812_SV_HV_clip_v2_UTM31N.tif

Additional Input Raster

Output Folder
E:\2022\KBAI\DLproces\DLtrainingsData\SV20200812_DLid

Input Feature Class Or Classified Raster Or Table
LVD_Annex1_Spec_habtype_2020119

Class Value Field
DLid

Buffer Radius
2

Input Mask Polygons

Image Format
TIFF format

Tile Size X
256

Tile Size Y
256

Stride X
64

Stride Y
64

Rotation Angle
0

Reference System
Map space

Output No Feature Tiles

Metadata Format
Classified Tiles

Train Deep Learning Model

Parameters Environments

Input Training Data
SV20200812

Output Model
SV20200812_UNet

Max Epochs
20

Model Parameters

Model Type
U-Net (Pixel classification)

Batch Size
8

Model Arguments

Name	Value
class_balancing	False
mixup	False
focal_loss	False
ignore_classes	0
chip_size	224
monitor	valid_loss

Advanced

Learning Rate

Backbone Model
ResNet-34

Pre-trained Model

Validation %
10

Stop when model stops improving

Freeze Model

Classify Pixels Using Deep Learning

Parameters Environments

Input Raster
20200812_SV_HV_clip_v2_UTM31N.tif

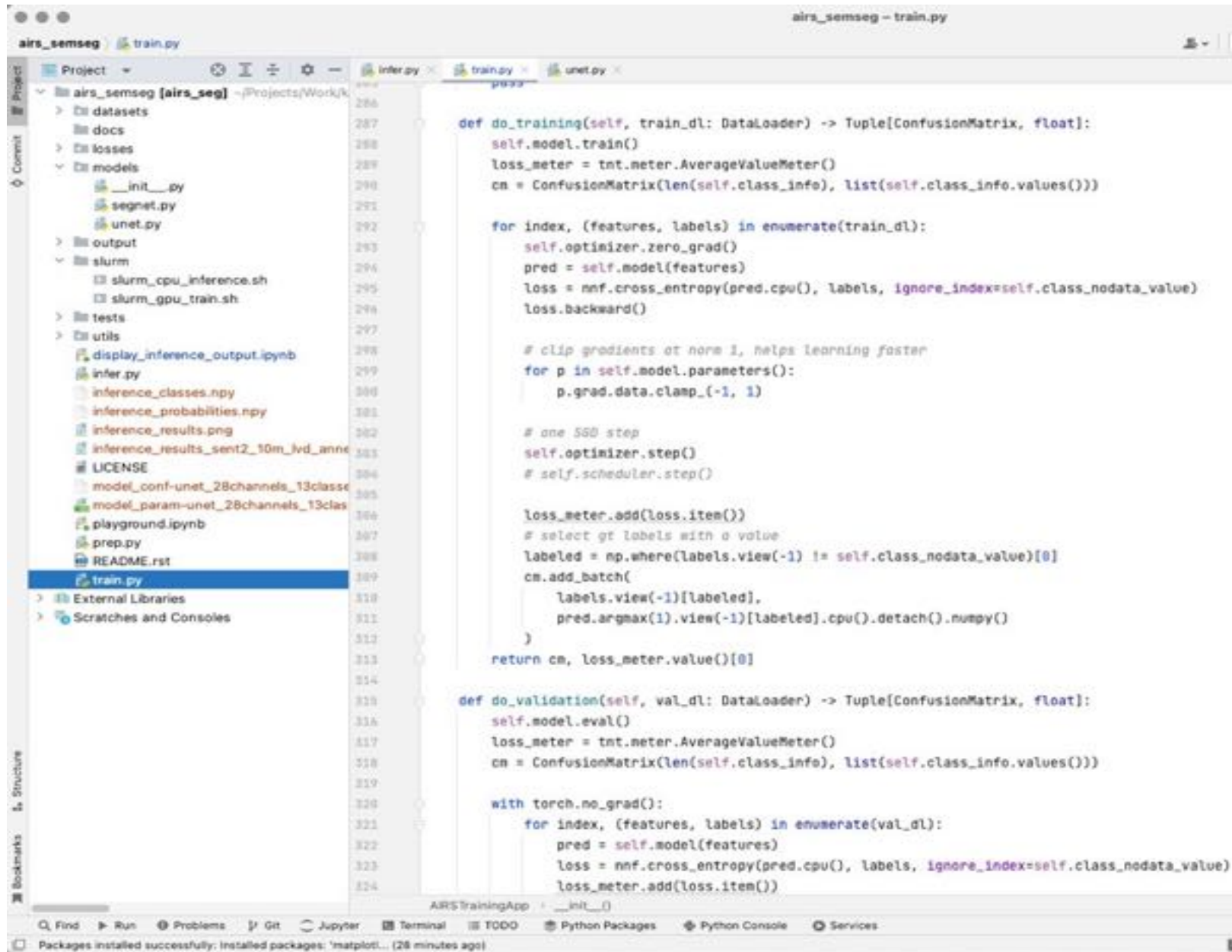
Output Classified Raster
20200812_SV_HV_clip_v2_UTM31N_UNetClass_DLid.tif

Model Definition
E:\2022\KBAI\DLproces\DLModels\SV20200812_UNet_DLid\SV20200812_U

Arguments

Name	Value
padding	56
batch_size	4
predict_background	True
tile_size	224

Method 2b Deep Learning in Pytorch with U-NET



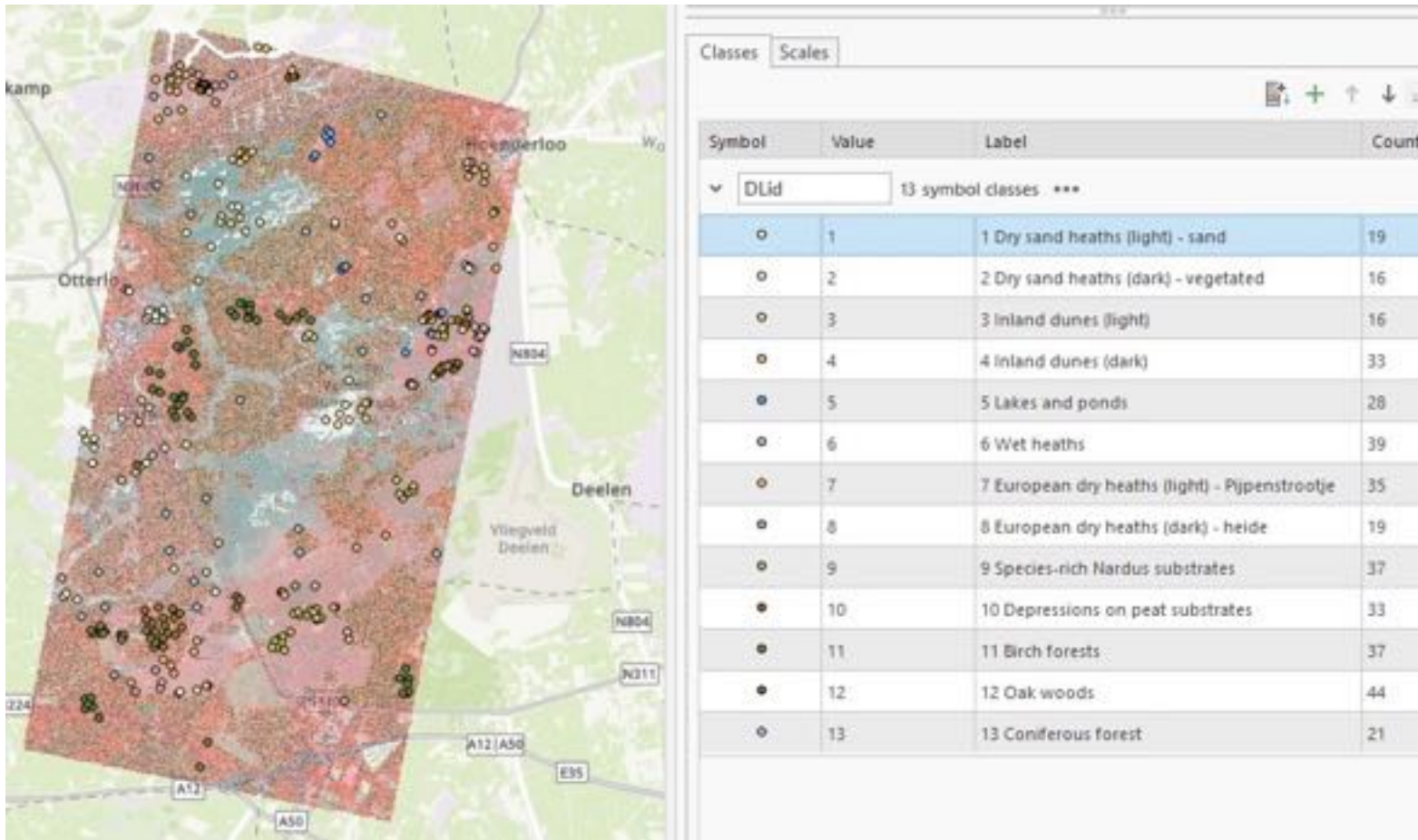
```
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

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)

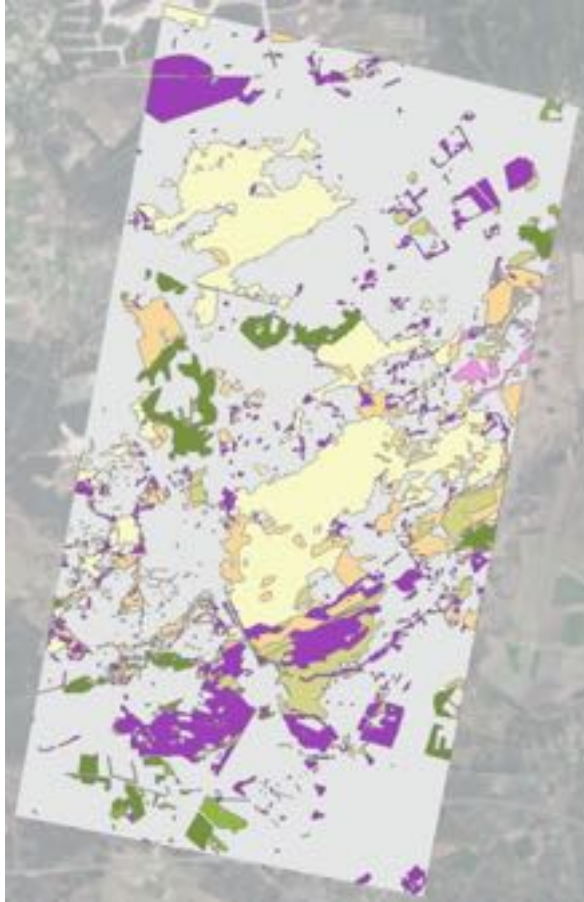


Habitat types Dry sands heaths (2310), Inland dunes (2330) and European dry heaths (4030) were divided into two subclasses each because for these three habitat types both light and dark appearances in the satellite image can be seen.

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.

U-NET deep learning technique

Groundtruth habitat map



Habitat map

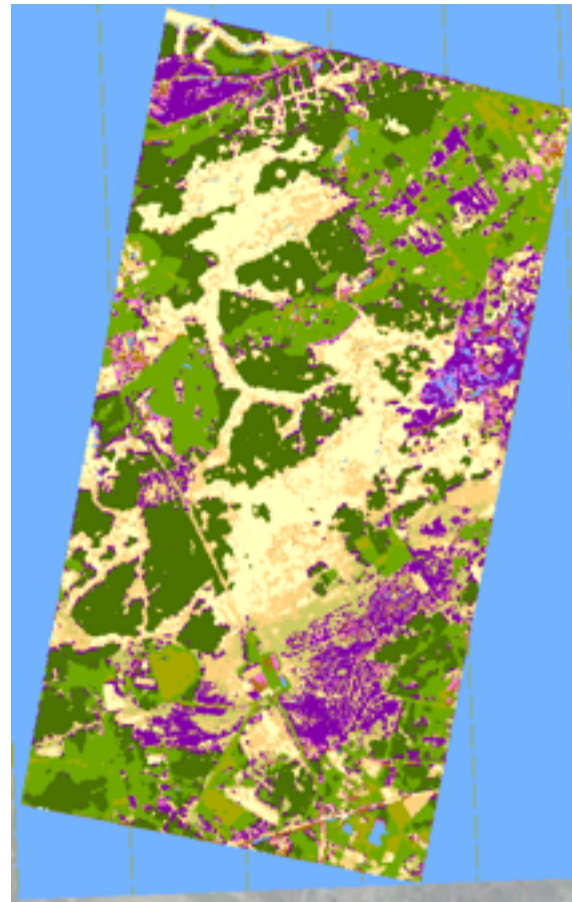
Result U-Net in ArcGIS Pro



Sentinel 2020 stack 7 images

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

Result U-Net in Pytorch

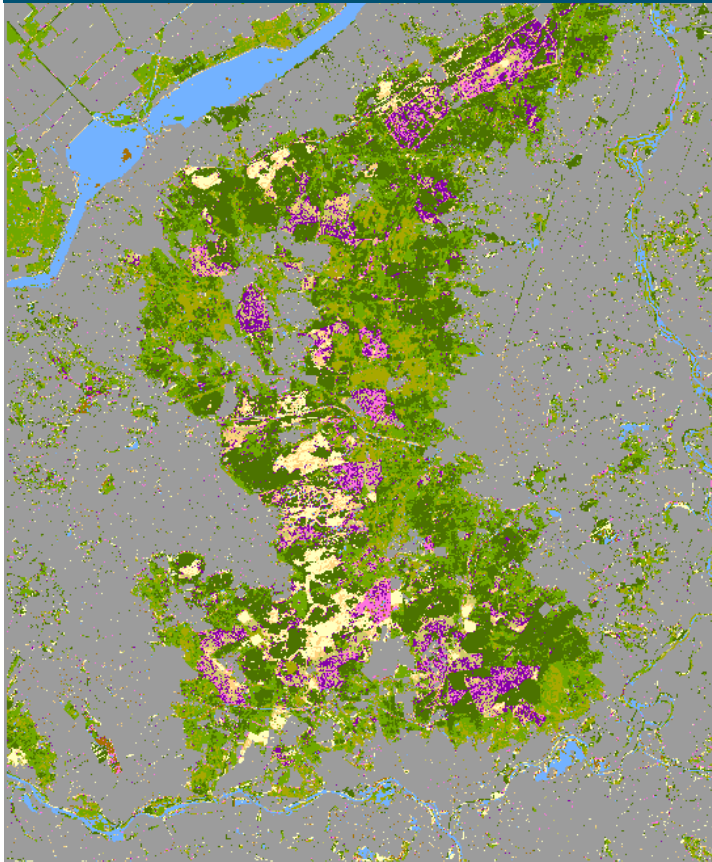


Sentinel 2020 stack 7 images
Classification in Pytorch

KB-DDHT-2 AI for Remote Sensing

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

Sentinel 2020 - 7 images



Deep learning
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



Validation

LDV training points	classification result											s	total	producer accuracy
	HABITATTYP	2310	2330	3160	4010	4030	6230	7150	9120	9190	forest			
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.

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

