Pixel-based mapping of open field and protected agriculture using constrained Sentinel-2 data

Daniele la Cecilia[a,1], Manu Tom[b,c], Christian Stamm[a] and Daniel Odermatt[b,c]

[a] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Environmental Chemistry, Überlandstrasse 133, 8600 Dübendorf, Switzerland.

[b] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Surface Waters - Research and Management, Überlandstrasse 133, 8600 Dübendorf, Switzerland

[c] University of Zurich, Department of Geography, Winterthurerstrasse 190, 8057, Zurich, Switzerland

Corresponding author: Dr. Daniele la Cecilia

Eawag

Department of Environmental Chemistry

Überlandstrasse 133, 8600 Dübendorf, Switzerland

daniele.lacecilia@eawag.ch

[1] Now at: Eawag, Swiss Federal Institute of Aquatic Science and Technology, Water Resources and Drinking Water, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland

Scenarios of spatial and temporal consistency

We formulate three scenarios to assess the implications on the output classification of a robustly trained model over time but not over space (scenario 1), over space but not over time (scenario 2) and over space on a particular season but not on other seasons. We assess such implications by means of five classification performance scores (i.e., Overall accuracy, precision, recall, F1-score and Intersection-Over-Union as in EQ 1 to 5 in the main manuscript) and we compile a confusion matrix for each scenario. For this analysis, we use the comprehensive case study of Switzerland; that is, we neglect labelled pixels sampled within the UTM grid square T30SWF of Spain because they were sampled for a single summer day in 2021, hence lacking temporal scope. In scenario 1, we train the model on the labelled pixels in the UTM grid square T32TLT and we validate it on the labelled pixels in T32TNT, and vice versa. In scenario 2, we train the model using the labelled pixels of both UTM grid squares on one year (2019) and we validate it using the labelled pixels on the other year (2021), and vice versa. In scenario 3, we split the dataset in 4 parts depending on the month of the year (MOY) (i.e, March to May, June to August, September to November and December to February). Then, a model is trained using the labelled pixels in one period of the year and validated on the labelled pixels in all other periods. This procedure is repeated for each different training period of the year.

In scenario 1, we assess spatial consistency. On the one hand, we calculate high performance scores relative to natural land covers (i.e., “Baresoil”, “Forest”, “Vegetated”, “Water” but not “Snow”) independently from the training region (Table 1). For “Snow”, the test using T32TLT as the training area is actually meaningless given that only 1 labelled pixel is in the dataset (Figure 4 in the main manuscript). On the other hand, for artificial land covers typical of protected agriculture, we calculate low performance scores (Table 1). This may be due to local decisions about plastic materials and management practices; the local characteristics should have been expected from the much different reflectance values measured in Band 1 between the two areas. Thus, spatial consistency for protected agriculture is not guaranteed.

In scenario 2, we assess temporal consistency. From the higher classification performance scores in this scenario (Table 2) as compared to the previous one, we see that yearly consistency is greater than the spatial one. The scores for the natural land covers are again higher than those of protected agriculture.

In scenario 3, we investigate the seasonality effect. Here, from the overall accuracies reported in Table 3, we conclude that there is a seasonal effect that affects the classification performance. In fact, the overall accuracies are higher along the diagonal of the table, when the model is trained and validated in the same season, but decrease away from the diagonal, when the model is trained in one season and validated in other ones.

These results corroborate a strong need for spatial variability in the dataset for robust land cover classification during generalisation across space. Temporal variability is less problematic as long as the model is trained in the same season of the product to be classified.

Table 1: Classification performance scores for scenario one, testing spatial consistency. For each score, the left column refers to the model trained on product T32TLT and validated on product T32TNT; vice versa for the right column. "Shadow" not reported because of the lack of labelled pixels for this class in the product T32TLT.

Class	*Precision*		*Recall*		*F1-Score*		*Intersection-Over-Union*	
"Baresoil"	0.95	0.95	0.99	0.92	0.97	0.93	0.95	0.88
"Forest"	0.9	0.68	0.74	0.64	0.81	0.66	0.68	0.49
"Glasshouse"	0.54	0.36	0.12	0.36	0.2	0.36	0.11	0.22
"Mulch_white"	0.13	0.86	0.81	0.07	0.22	0.13	0.13	0.07
"Snow"	0.09	0.99	0.99	0.86	0.17	0.92	0.09	0.86
"Tunnel"	0.69	0.62	0.71	0.92	0.7	0.74	0.54	0.59
"Vegetated"	0.94	0.93	0.98	0.93	0.96	0.93	0.92	0.87
"Water"	0.99	0.88	0.91	0.99	0.95	0.94	0.91	0.88

Table 2: Classification performance scores for scenario two, testing yearly consistency. For each score, the left column refers to the model trained in year 2019 and validated in year 2021; vice versa for the right column.

Class	*Precision*		*Recall*		*F1-Score*		*Intersection-Over-Union*	
"Baresoil"	0.97	0.94	0.94	0.99	0.95	0.96	0.91	0.93
"Forest"	0.85	0.77	0.57	0.87	0.68	0.82	0.52	0.69
"Glasshouse"	0.34	0.86	0.64	0.3	0.44	0.44	0.29	0.29
"Mulch_white"	0.64	0.46	0.16	0.71	0.26	0.56	0.15	0.39
"Shadow"	0.76	0.99	0.95	0.92	0.84	0.96	0.73	0.92
"Snow"	0.99	0.99	0.94	0.97	0.97	0.98	0.94	0.97
"Tunnel"	0.78	0.67	0.81	0.87	0.79	0.76	0.66	0.61
"Vegetated"	0.92	0.94	0.97	0.88	0.94	0.91	0.9	0.83
"Water"	0.99	0.85	0.97	0.58	0.98	0.69	0.96	0.52

Table 3: Overall classification accuracy for scenario three, testing seasonal consistency. For the cells along the diagonal, the model is trained on 80% of the dataset and validated on the remaining 20% for the period of the year shown either in the columns or the rows. In the cells outside the diagonal, the model is trained in the period of the year defined in the columns and validated in period of the year reported in the rows.

Period of the year	March to May	June to August	September to November	December to February
March to May	0.95	0.85	0.87	0.83
June to August	0.71	0.90	0.80	0.77
September to November	0.75	0.84	0.94	0.84
December to February	0.83	0.86	0.89	0.91