Supplementary Material

Qualitative analysis For a more qualitative evaluation of our PSE+TAE architecture, we provide its confusion matrix on the test set on Figure 2. We note that many of the errors are misclassification as *Meadows*, the most represented class in our dataset. Additionally, the model struggles to discriminate between *Winter Durum Wheat* and *Winter Cereal*, likely due to their similar phenology.

We also show a visual representation of our model's prediction errors compared to those of a CNN+TAE architecture on Figure 1. While the PSE+TAE without f corrects some errors made by the CNN+TAE (the two parcels marked with (1) on Figure 1a) , it produces new errors ((2) on Figure 1b) . The geometric features in the full PSE+TAE architecture allow to correctly classify the latter and yield a wrong classification only for the two parcels (3) (Figure 1c) that belong to hard classes (*Winter Durum Wheat* and *Leguminous Fodder*) and where incorrectly classified by all models.

Processing time profiling We provide a breakdown of the processing times during training for the different architectures in Table 1. The average time per batch is decomposed into data loading time, forward pass and gradient back-propagation. We can see that the processing time is dominated by the loading time except for the Transformer which processes pre-computed means.

Architecture hyperparameters We show the exact configuration of our PSE+TAE architecture on Table 1 hyperparameters, as well as those of the different competing methods in Table 2.

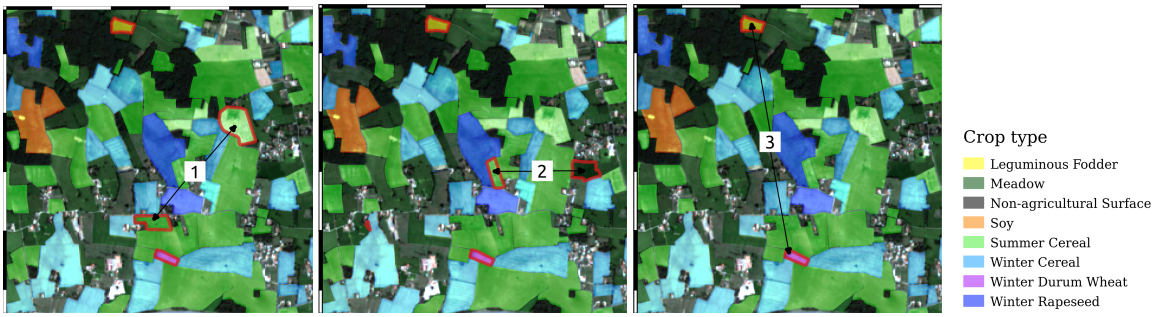
Dataset composition Lastly, we show the class breakdown of our dataset on Figure 3 on a semi-logarithmic scale. The dataset is highly unbalanced: half of the samples (around $100,000$ parcels) belong to the *Meadow* class. The next most prominent classes are *Winter Cereal*, *Summer Cereal*, *Grapevine* with more than $10,000$ parcels each. Lastly, many classes are only represented by a few hundred samples. The ability of a model to learn from these few samples is thus critical to achieve a satisfactory performance across the nomenclature.

Time in ms/batch	Total	Loading	Forward	Backward
PSE+TAE (ours)	**107**	**85**	11	**11**
CNN+TempCNN	381	365	**4**	12
CNN+GRU	437	365	14	58
Transformer	8	1	2	5
ConvLSTM	530	365	61	104

Table 1: Comparison of processing time for different methods for batches of 128 parcels.

	Number of parameters
CNN+GRU	144 204
• 3×3 convolutions: 32, 32, 64 kernels • Global average pooling • Fully connected layer: 128 neurons • Hidden state size: 130	
CNN+TempCNN	156 788
• 3×3 convolutions: 32, 32, 64 kernels • Global average pooling • Fully connected layer: 64 neurons • Temporal convolutions: 32, 32, 64 kernels of size 3 • Flatten layer	
Transformer	178 504
• $d_k = 32$, $d_v = 64$, $d_{model} = 128$, $d_{inner} = 256$ • $n_{head} = 4$, $n_{layer} = 1$	
ConvLSTM	178 356
• Hidden feature maps: 64	
RF	
• Number of trees: 100	

Table 2: Hyperparameters of the competing architectures. For all models we use the same values for the decoder MLP_4.



(a) CNN+TAE (b) PSE+TAE *no f* (c) PSE+TAE

Figure 1: Example of test-errors of three architectures on a sub-region of the dataset. The images consist in the RGB channels of a single Sentinel-2 observation overlayed with a color-coded representation of the different parcels' crop types. Those parcels that were wrongly classified by the model are highlighted with a solid red stroke. The scale is given by the 500 meter zebra strips. We compare the errors of the CNN+TAE (a), the PSE+TAE *without geometric features* (b), and the complete PSE+TAE (c).

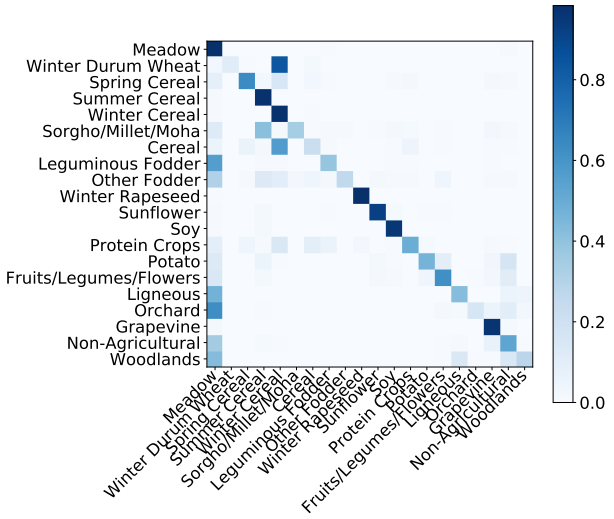


Figure 2: Confusion matrix for our PSE+TAE architecture on the AOI. The color represents the number of parcels, expressed relatively to the total population of the class they belong to.

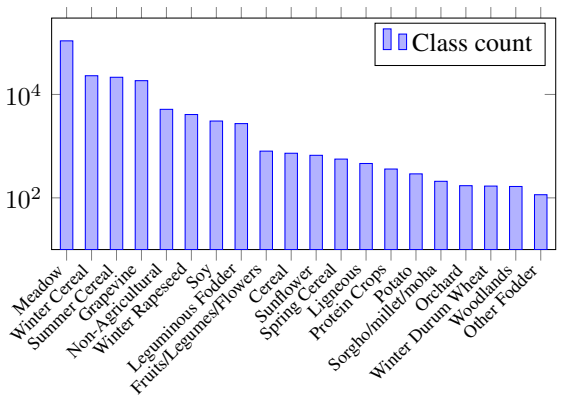


Figure 3: Class repartition in the AOI.