High Carbon Stock Estimation with Sentinel-2 and Deep Learning Methods

Semester Project - Final Report

Olivier Dietrich

M.Sc. Robotics, System & Control

Supervised by

M.Sc. Nikolai Kalischek

M.Sc. Nico Lang

Dr.-Ing. Jan Dirk Wegner

June 8, 2020

Contents

1 Introduction

This report presents the results of my semester project entitled *High Carbon Stock Estimation with Sentinel-2 and Deep Learning Methods*. This project is an integral part of my M.Sc. in Robotics, System & Control and is worth 8 ECTS. It was realized during the Spring semester 2020. More accurately, it officially started on February the 24th and ended on June the 5th. It was supervised by M.Sc. Nikolai Kalischek, M.Sc. Nico Lang and Dr.-Ing. Jan Dirk Wegner, all from the EcoVision Lab at ETH Zürich.

My main motivation in choosing this topic was the combination of two elements to which I attach great importance. First, I have been focusing my studies towards the fascinating field of deep learning and computer vision, and I was looking for a project to gain hands-on experience on it. Second, my professional ambition is to apply the knowledge and skills acquired in class to solve real-world issues and have a positive impact both socially and environmentally on our society. This project perfectly combined these two elements, since state-of-the-art technology is used to help prevent deforestation.

In the following pages, I will detail the different steps undertaken in this project. Section 2 will briefly recap what the High Carbon Stock Approach is and how it works. Section 3 will introduce the technical aspects of this project in a more detailed way. Then, the data used will be listed in Section 4, while the deep learning model chosen will be explained in Section 5. Then, Section 6 and 7 will be about the results obtained. At last, Section 8 will discuss the validity of the results obtained.

2 High Carbon Stock Approach

According to the FAO[1], the total forest area is 4.06 billion hectares, which cover approximately 30 percent of the world's land. However, forest area accounted for 5 billion hectares in the beginning of the 20th century. Since 2015, the rate of deforestation is estimated at 10 million hectares per year. Although it has been decreasing[2], this rate is still alarmingly high and entails disastrous consequences such as loss of biodiversity, climate change, soil erosion or destruction of livelihoods for local communities.

Agricultural expansion continues to be the main driver of deforestation, and large-scale commercial installations, mostly cattle ranching and single-crop farming such as soy bean or oil palm, accounted alone for 40 percent of tropical deforestation between 2000 and 2010. As part of UN's Sustainable Development Goals (SDGs), forests sustainable management are now more than ever on the agenda of *most* governments worldwide, and several methods have been proposed to tackle this issue. One of them is called the **High Carbon Stock (HCS) Approach**.

The HCS Approach is a methodology that distinguishes forest areas for protection from degraded lands with low carbon and biodiversity values that may be developed. Promoted by Greenpeace and supported by big food companies such as Barry Callebaut or Unilever, it gives a framework to prevent deforestation while ensuring the rights and livelihoods of local peoples are respected. In short, the HCSA classifies forests into six classes, shown on Figure 1, based on several criteria such as canopy closure or metric tons of carbon per hectare (Ct/ha) (cf Table 1). The four classes above the HCS threshold (*i.e.* HDF, MDF, LDF and YRF) should be protected whereas *Scrubs* (S) and *Open Land* (OL) could be used for potential further development.

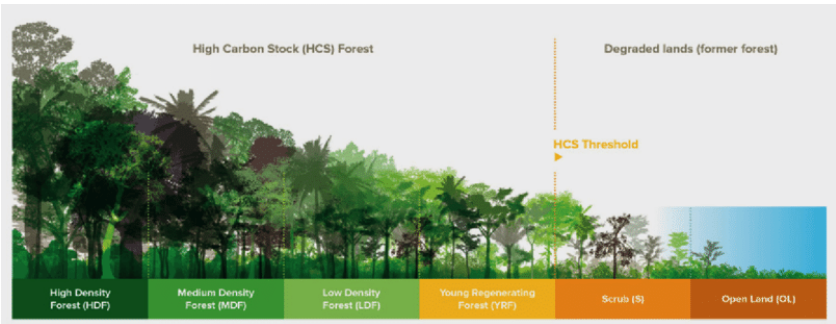


Figure 1: Vegetation stratification according to HCSA

[1] http://www.fao.org/state-of-forests/2020/en/
[2] It was estimated at 16 million hectares per year in the 1990s

HCS Classes	OL	S	YRF	LDF	MDF	HDF
Carbon t/ha	0-15	15-35	35-75	75-90	90-150	>150

Table 1: Correspondences between metric tons of carbon per hectare (Ct/ha) and HCS classes

This methodology strongly relies on accurate remote sensing tools. While direct field measurements can be conducted and result in accurate classification, they should be kept for validation as they are tedious to obtain, give sparse maps and cannot be scaled to large areas. Drones equipped with LiDAR can be used to map canopy height at larger scale[3] with great accuracy, but it is an expensive process whose range is still limited to small regions. Moreover, deforestation dynamics adds the challenge of keeping data up-to-date, which discards drones from optimal solutions. The most promising approach seems therefore to leverage the use of satellite images, which give both high temporal and spatial resolutions. As detailed in section 4.1, ESA satellites Sentinel-2 produce publicly available multi-spectral images of the whole world with a revisit time of five days and are particularly well suited for this task.

3 The Project

This work is part of the larger *Automated Large-scale High Carbon Stock Estimation from Space* project, a joint ETH Zürich-Barry Callebaut project led by the EcoVision Lab. It aims to leverage the use of available space data with deep learning data-driven techniques to build an objective, highly automated tool to guide sustainable agribusiness. Recently, Lang et al.[3] predicted canopy height from Sentinel-2 images at a 10 m spatial resolution in Gabon and Switzerland. Similarly, my goal is to regress a HCS map for Sekadau and Sangau, two regions in Western Kalimatan, Borneo, Indonesia. These regions, suggested by Greenpeace, encapsulate a large spectrum of vegetation and are threatened by a fast deforestation rate due to the palm oil industry.

A major drawback with supervised deep learning methods is their reliance on the existence of quality labelled dataset that can be used to train the models on. For my work, the initial goal was to cross-validate NASA's GEDI data with some samples of field measurements to get a reliable groundtruth in order to regress a HCS map from Sentinel-2 data. GEDI is a LiDAR instrument boarded on the International Space Station (ISS) specifically designed to map ecosystems. In fact, it is the LiDAR with the highest resolu-

[3]Canopy height is strongly correlated to carbon stock.

tion and densest sampling ever put into orbit[4]. Unfortunately, due to technical reasons, these data were never made available as initially planned. The only other groundtruth available was a biomass map from ESA. It is made of data from 2017 and has a lower resolution than Sentinel-2 data. Furthermore, the field measurements in our possession did not match the ESA map and had to be discarded as well. Consequently, this project had to be largely reconsidered from its initial goal. With the very limited amount of data at our disposal, it was decided to change the approach and do first a land cover classification using handmade labels instead. The motivation was that this land cover map could then be used to filter the ESA biomass map to obtain a more reliable and up-to-date biomass map. Also, due to the extra complications that came along the lack of reliable groundtruth, it was decided to work on one Sentinel-2 tile only and focus the time available on improving and refining the deep learning models instead of handling several tiles with multiple locations or dates.

4 Data

4.1 Sentinel-2

All the data collected by the two satellites from the Sentinel-2 mission are easily accessible via the Copernicus Open Access Hub[5]. Each Sentinel-2 tile spans an area of 100x100km^2 and consists in 13 spectral bands with resolution going from 10m for the red, green, blue and near-infrared (NIR) bands to 60m, as shown in Figure 2. The tiles are proposed at different levels of processing. The level 2-A was chosen as it includes atmospheric corrections. However, the cirrus band B10 is not available at this level and the final data contains 12 layers only.

Clouds are a consistent problem while working with satellite images. Sentinel-2 data include a useful cloud cover map that can be used to filter cloud coverage while choosing the data to work with. The final tile selected comes from August 2019 and contains less than one percent of clouds[6]. The RGB version of it is shown on Figure 3.

[4]https://gedi.umd.edu/instrument/instrument-overview/

[5]https://scihub.copernicus.eu

[6]URL to download the tile: https://scihub.copernicus.eu/dhus/odata/v1/Products('bf1a5ee4-1088-487e-838b-b42e174f91d4')/$value

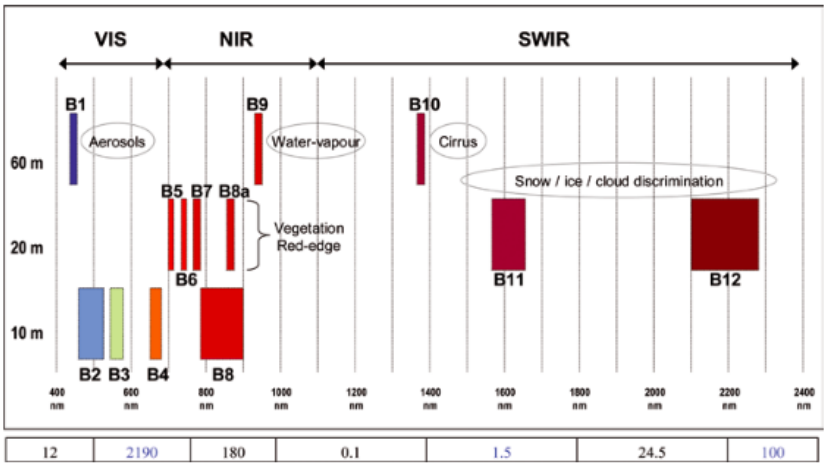


Figure 2: All 13 spectral bands vs spatial resolution

Figure 3: Sentinel-2 tile used for the project 100x100km^2

4.2 Handmade Labels

As mentioned earlier, an important part of this project became the land cover classification, for which labels were needed. Labels were made by hand using open-source software QGIS. The five following classes were chosen to match HCS approach:

- Vegetation (native)
- Water Bodies
- Built-up Areas, Settlements
- Concessions, Palm Oil Plantations
- Open Land

where the bullet color corresponds to the color code used throughout the project. Labelling satellite images by hand is a tedious process prone to human errors. In particular, distinguishing palm trees from native vegetation can be delicate. As shown in Appendix B, I navigated between the Sentinel-2 image and Google Maps data to cross validate my labels, and I refined them several times as I was moving forward with my project. Nevertheless, these labels still contain some inaccuracies and this should be remembered while measuring the model performance, as explained more in details in section 6.2. The entire labelled tile can be found in Appendix A, but some examples of these labels are shown on Figure 4 below. In total, more than 1150 polygons were made.

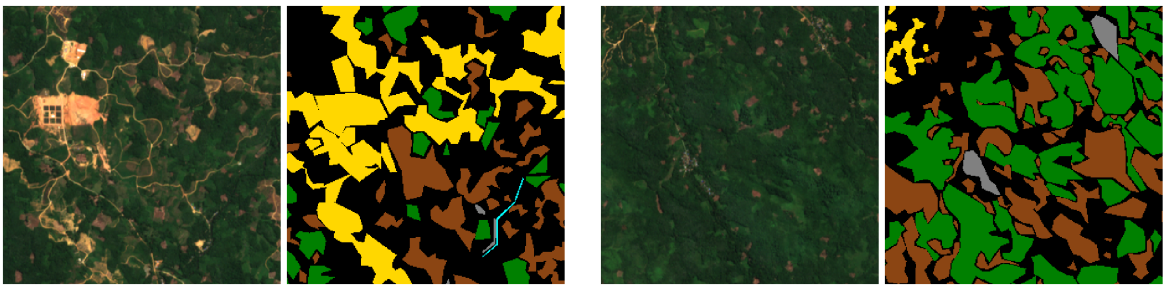
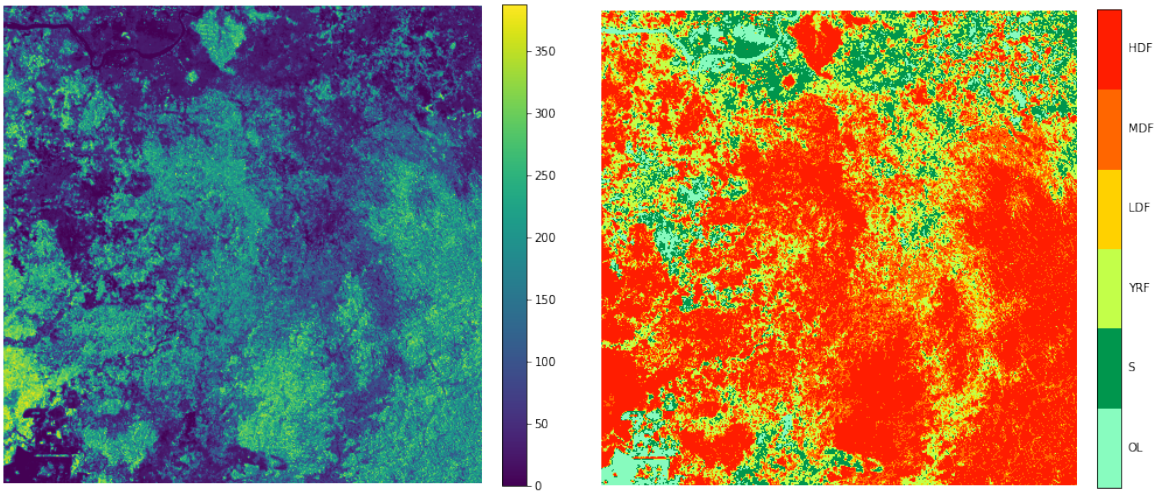


Figure 4: Examples of hand-labelled areas (300x300m^2). Black pixels are non-labelled.

4.3 ESA biomass map

The 2017 ESA biomass map[7] is the only groundtruth at our disposal for HCS values. It spans the same area as the Sentinel-2 tile but has a much lower ground resolution of 100x100m^2 per pixel. The maximum HCS value for this tile is 388 Ct/ha and the mean is 128.3 Ct/ha.

The map was interpolated to the desired 10m resolution using 3rd order spline interpolation and is shown on Figure 5. It has continuous HCS values but, in order to ease visualization, a thresholded version with values from Table 1 is also plotted to show the distribution of HCS classes.



(a) Continous spectrum

(b) HCS classes

Figure 5: Interpolated ESA biomass map for the Sentinel-2 tile.

5 Model

In the last few years, deep learning using Convolutional Neural Networks (CNNs) has become the predominant tool for many computer vision tasks due to the CNNs' ability to extract patterns from images. After the success of AlexNet[2] at the ImageNet Large Scale Visual Recognition Challenge in 2012, these types of architectures have been widely studied and many different networks have been proposed. Lang et al.[3] used an adapted version of the Xception model[1] to predict canopy heights in Switzerland and Gabon

[7]ESA Biomass Climate Change Initiative: Global datasets of forest above-ground biomass for the year 2017 (https://catalogue.ceda.ac.uk/uuid/bedc59f37c9545c981a839eb552e4084)

from Sentinel-2 images. The resulting maps show a very low mean absolute error of only 1.7m for Switzerland and 4.3m for Gabon. Since both my task and the data are similar, I will use the same model for my project. The model was thoroughly detailed by Lang et al. and consists in one entry block followed by a series of eight identical *separable convolution* (SepConv) blocks, as shown in Figure 6.

The entry block is made of three pointwise convolutional layers that increase progressively the channel depth from 12 to 728 channels. Each pointwise convolution is followed by a batch normalisation (BatchNorm) and goes through a rectified linear units (ReLU). A skip connection bypasses the three pointwise convolutional layers and is added to the resulting activation maps right before the last ReLU activation. For dimensional reasons, the skip connection also goes through a linear convolutional layer and a BatchNorm to increase its depth to the required 728 channels.

Each of the 8 SepConv blocks is made of two depthwise separable 3x3 convolution layer, preceded by a ReLU activation function and followed by a BatchNorm layer.

A final pointwise convolution combines the activation maps into a specific value per pixel. In the case of the land cover classification, the model predicts five probabilities for each pixel ($N_{out} = 5$), corresponding to the five land cover classes, and the maximal one is kept as prediction. For the HCS regression task, the model returns directly the HCS value regressed for each pixel ($N_{out} = 1$).

Overall, this network with 8 SepConv blocks has 8'847'174 trainable parameters. It should be observed that the size of the activation maps are never up- nor down-sampled. The input of each 3x3 convolution is always padded to keep the same size.

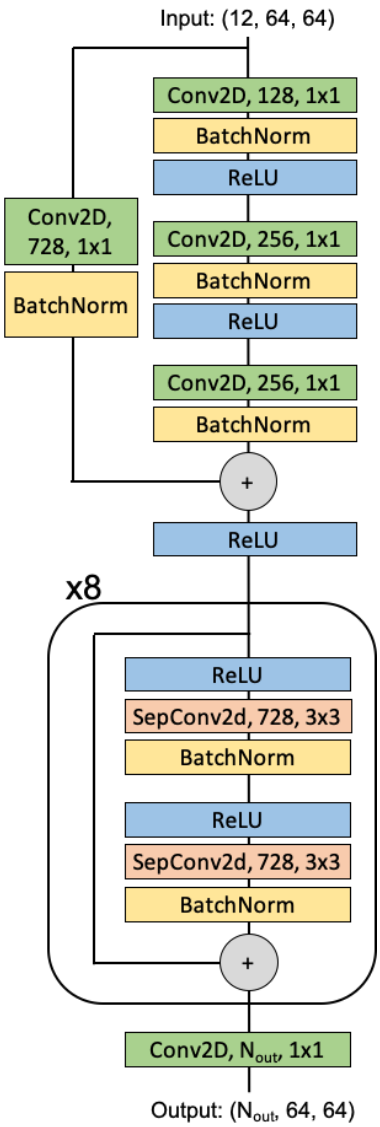


Figure 6: Model overview

The other parameters in the learning process such as the loss, the optimizer, or the training and validation areas depend on the task and will be described in the following two sections.

6 Land Cover Classification

In this first part of the project, a land cover classification will be computed from hand-made labels.

6.1 Training

To classify each pixel of our Sentinel-2 tile into one of the 5 classes stated in section 4.2, the Sentinel-2 tile was first splitted into smaller patches of size 64x64. The top-right part of the tile was kept for validation, as shown in Figure 7. To ensure, each patch is relevant for the training, all patches that have less than 40% pixels labelled have been discarded. This results in 405 patches for training and 116 for validation.

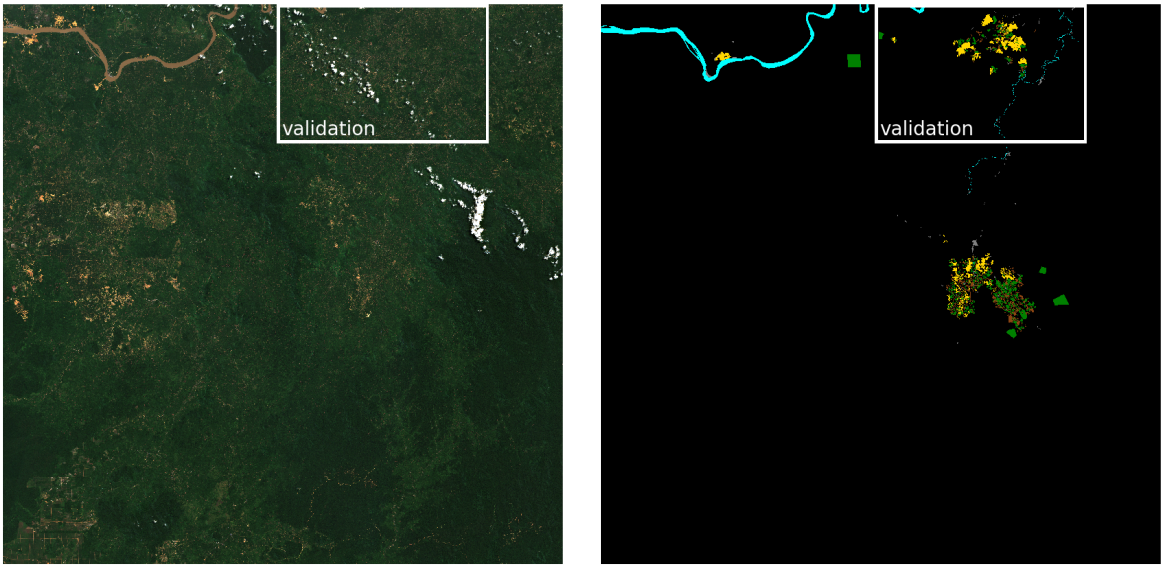


Figure 7: Validation area used during land-cover classification

The classes are strongly unbalanced between each other and between the training and validation areas (see Figure 8). In order to overcome this issue, the weighted cross-entropy (WCE) loss has been used, which allows to measure the classification performance of a model when the classes are not uniformly distributed. The smaller the loss gets, the better the model is. Formally:

$$Loss_{WCE} = -\frac{1}{N}\sum_{o=1}^{N}\sum_{c=1}^{C} w_c \cdot y_{o,c} \cdot log(p_{o,c}) \tag{1}$$

where N is the number of pixels, C the number of classes (*i.e.* 5), w_c the weight of each class, $y_{o,c}$ a binary indicator if the class c is the correct indicator for observation o and

$p_{o,c}$ the probability predicted that observation o belongs to class c. The weights w_c are found with:

$$w_c = \frac{1}{n_c \cdot \sum_{i=1}^{C} \frac{1}{n_i}}$$

where n_c is the number of pixels belonging to class c. The smaller the loss means the better the model. Moreover, the already-implemented *nn.CrossEntropyLoss()* function from the library torch.nn has a handy additional parameter *ignore_index* that allows to ignore the unlabelled pixels in the loss calculation and avoid the need of a custom loss function.

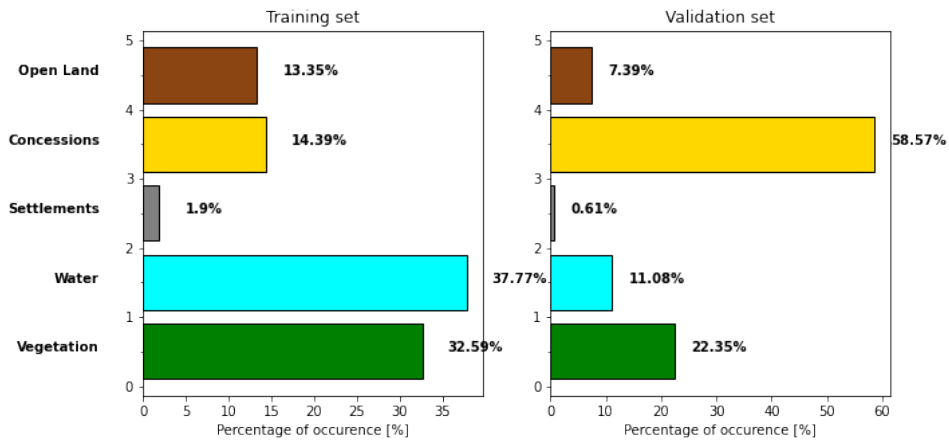


Figure 8: Distribution of each class in the training and validation area

The data were augmented using random horizontal and vertical flips. After tuning, the learning rate was set to 10^{-5}, the batch size to 5 and the optimizer used was adam without any weight regularization. The model saturates quite rapidly, and even overfits as the validation loss increases after about 200 epochs. as shown in Figure 9. Each epoch takes about 50 seconds on the local GPU.[8]

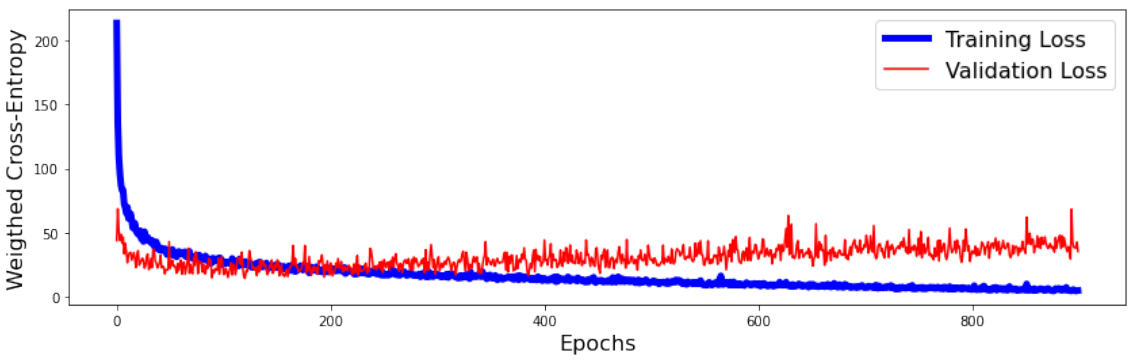


Figure 9: Training and validation loss for the land cover classification

[8]GeForce GTX Titan X

6.2 Results

The land cover predictions for the full tile is shown on Figure 10. This is the result after 300 epochs, which was visually found to be the best.

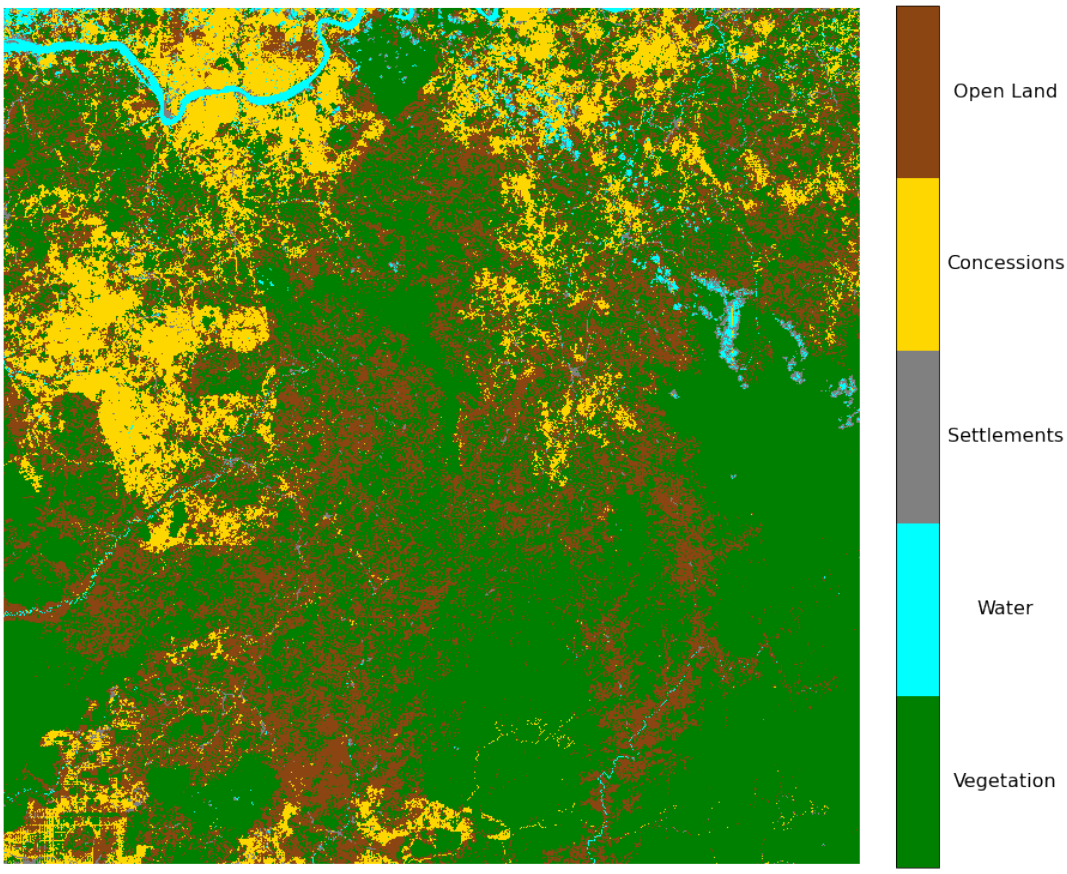


Figure 10: Final land cover predictions after 300 epochs

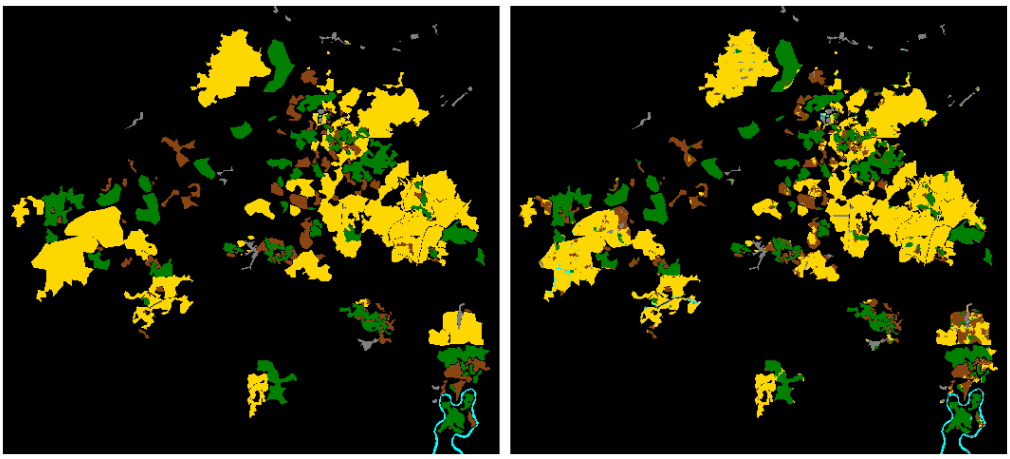


Figure 11: Visual comparaison between groundtruth (left) and prediction (right) for validation area.

The predictions are compared to the validation labels visually, as shown on Figure 11, and analytically, through the confusion matrix shown on Figure 12. Again, one should be cautious with these results since the labels have been made by hand and, despite having been refined several times, are still prone to human mistakes and inaccuracies.

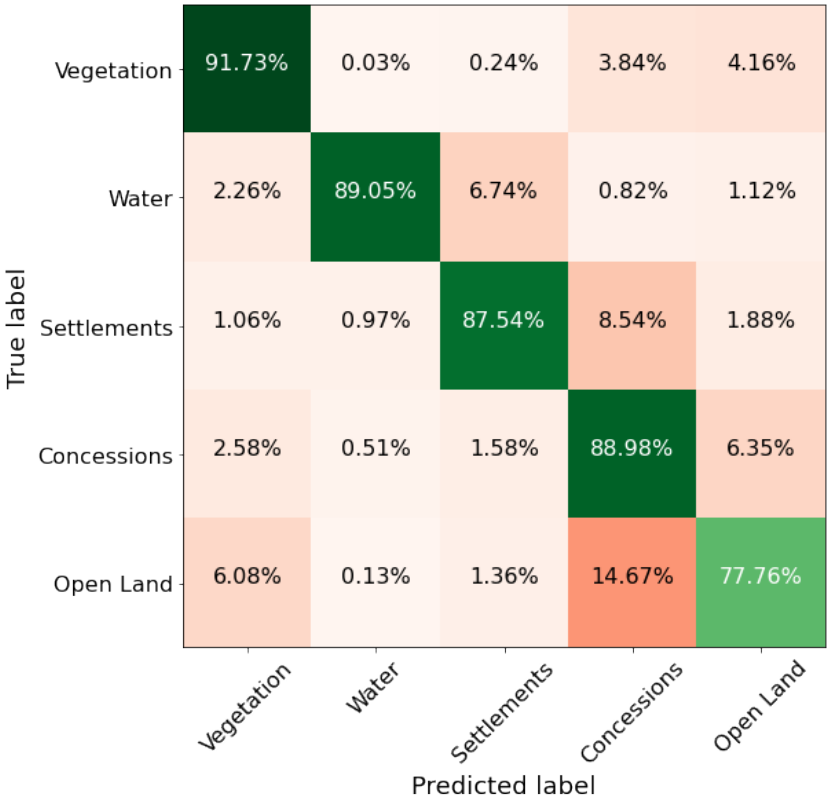


Figure 12: Confusion matrix on the validation area. Total accuracy is 88.23%.

A few observations can be made. First, according to the HCS Approach Toolkit[9], the accuracy of a preliminary land cover classification must reach 70% at least and the one for the final land cover classification 80%. This model returns an overall accuracy of **88.23%**. Then, it classifies correctly 91.73% of the native vegetation labelled, which is the most important class since it is the one that needs to be protected. Its major weakness seems to be the open land class, as it fails to recognize almost a quarter of it. This, however, does not come as a surprise. The distinction between both classes has posed some challenges since the beginning. Indeed, it is not always clear which class certain areas should belong to. It happens for instance that some areas clearly belong to some palm oil concessions, but have also been cleared recently, which makes it hard to label. Figure 13 is a zoom on the bottom-right part of Figure 11, where a big area labelled as concession has been *wrongly* classed by the network. On one hand, all these

[9]HCS Approach Toolkit, Module 4, p.15-16

rectangle shapes induce that this area is a concession, and it has thus been labelled as such. On the other hand however, some of the fields may not be cultivated yet and thus may well be open land or even native vegetation. In Appendix B, I show the Google Maps satellite image of the same area, which is of higher resolution but of different date. Even with this higher resolution, it is hard to know whether some of these rectangle fields should have been labelled as open land or not.

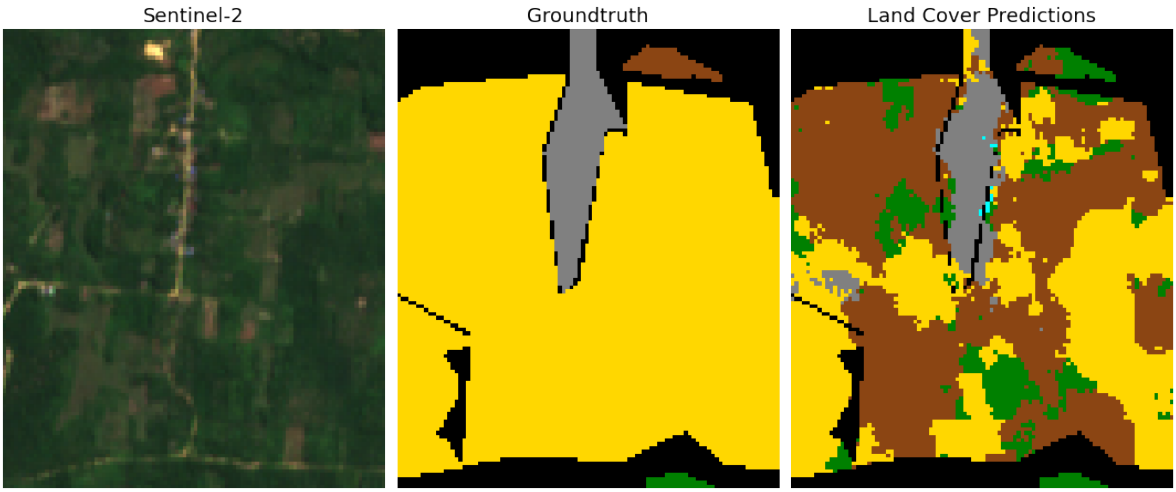


Figure 13: Zoom on an area misclassified.

At last, one can zoom on a national park to make sure that the model predicts it correctly. Figures 14 shows a national park selected on the tile. The groundtruth was made separately from the other polygons shown before. It can be seen that it is almost 100% correctly predicted with the obvious exception of the clouds. It is also interesting to note that clouds and clouds shadows seem to be predicted as water bodies and built-up areas respectively. Since Sentinel-2 data come with a cloud mask, one can easily filter them out in a post-processing step.

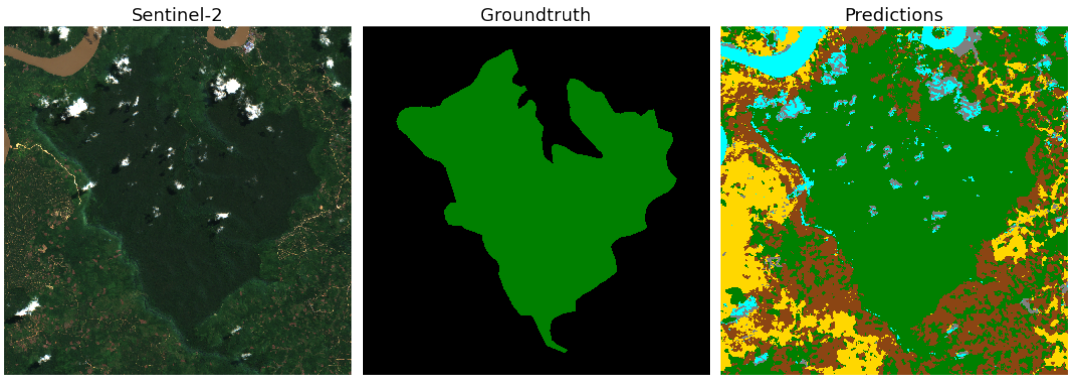


Figure 14: Zoom on 1st national park selected

7 HCS regression

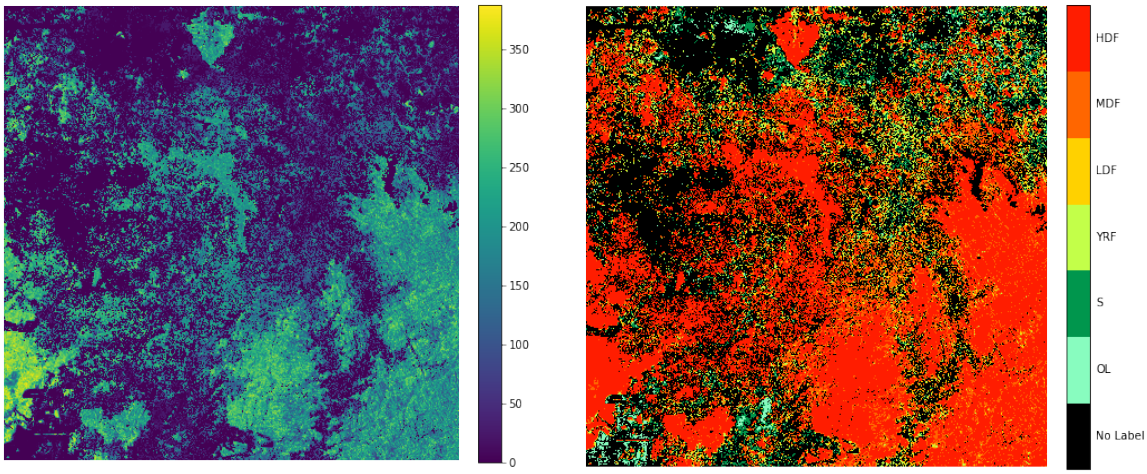
In this second part, the land cover classification will be used to regress a new HCS map.

7.1 New Labels

In order to conduct a regression and predict a new HCS value for each pixel, the land cover labels predicted in section 6 were combined with the ESA map following this *recipe*:

- From the land cover classification, keep only the pixels labelled as *Vegetation* or *Open Land*. The other categories should not have HCS values.
- For these pixels, impute the HCS value from the ESA map. If value does not match the class (*i.e.* *Open Land* above 35 or *Vegetation* below 35), discard.
- Using the mask provided by Sentinel-2, discard all pixels that belong to clouds.

The new labels can be seen in Figure 15. As in section 4.3, the map is shown with both continuous spectrum and HCS classes. From now on, only the HCS classes will be shown.



(a) Continous spectrum (b) HCS classes

Figure 15: New labels for HCS regression task from combination between land cover classification and ESA biomass map

Technically, this HCS map used as new groundtruth for regression could already be the final product that we are looking for. However, one should keep in mind that the ESA map is outdated and has been upsampled. Therefore, we hope that the model will be able to *correct* these inaccuracies when predicting the HCS values directly from the Sentinel-2 data.

7.2 Training

Ideally, the whole tile should be used to train with only a small part kept as validation. However, the time required to train the model on such a big dataset with our sole GPU would be way too large. Also, at the time of this task (mid-May 2020), ETH's Leonhard cluster was shut down for security reasons and could not be of any help. Consequently, smaller training and validation areas were defined on the tile. They were chosen as shown in Figure 17. The training n°2 area was added to ensure that there was a sufficient amount of *Open Land* in the training set. Figure 16 shows the distribution of HCS values for each area. Also, this kind of model suffers from boundary effects due to the padding needed for the convolution. In order to mitigate this effect, the patches of 64x64m^2 were cut in such a way that they overlap each other by half the size for the training dataset (*i.e.* the sliding window only moves of 32 pixels between each patch). As for the land cover classification, the patches with less than 40% of pixels labelled were discarded. This results in 5653 64x64 patches for training and 951 for validation.

The same Xception model described in Section 5 is used for the regression. This time however, the number of channels predicted is changed to 1 since a regression is being performed. The loss is a custom Mean-Square Error (MSE) which simply ignores the non-labelled pixels through a mask. Formally:

$$Loss_{MSE} = \frac{1}{N_{\mathbb{1}_p}} \sum_{p=1}^{N} \mathbb{1}_p \cdot (y_p - \hat{y}_p)^2 \tag{2}$$

where $\mathbb{1}_p$ is the indicator function used as a mask (*i.e.* 1 if pixel p labelled, 0 otherwise), y_p is the groundtruth HCS value for pixel p and $\hat{y}_p$ is the predicted HCS value. N is the total number of pixels and $N_{\mathbb{1}_p}$ is the number of pixels labelled in the groundtruth.

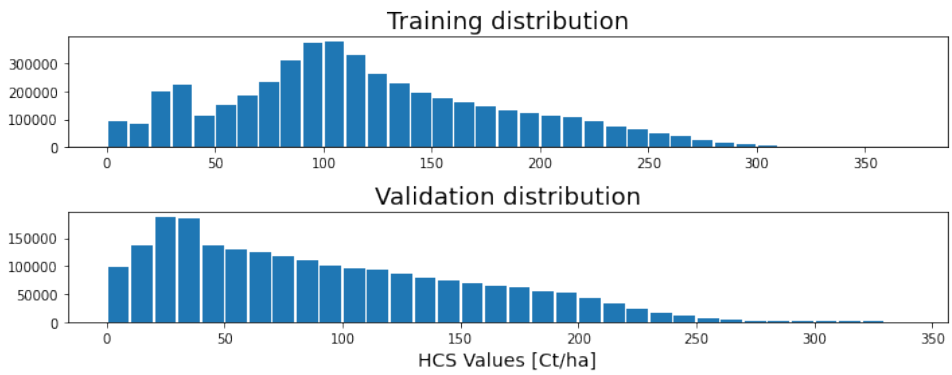


Figure 16: HCS values disctribution for both training and validation areas

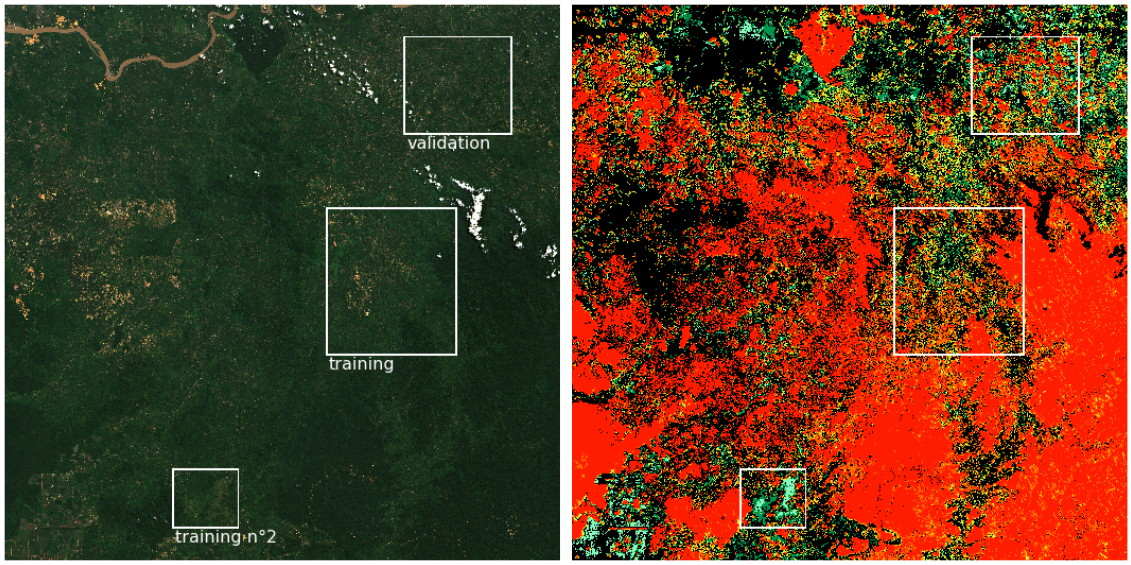


Figure 17: Training and validation areas for HCS regression task

Unfortunately, due to time constraints, the training configuration and hyperparameters could not be tuned in an optimal manner. In particular, the validation loss still does not behave as expected, as shown in Figure 18, and the attempts to solve it have not been successful yet. Nevertheless, the configuration which gave these losses and the results obtained are still presented in the following paragraphs.

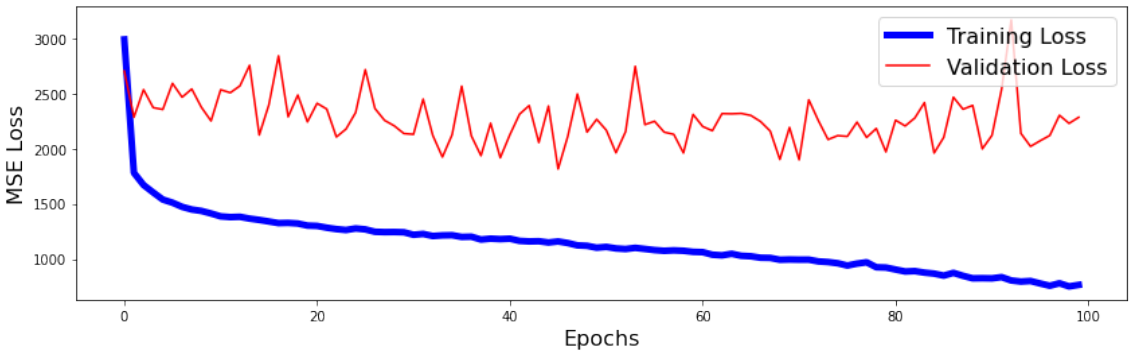


Figure 18: Training and validation loss for the HCS regression.

The data were augmented using random horizontal and vertical flips, and random rotation with an angle up to 15°. The learning rate and batch size were not modified and were kept at respectively 10^{-5} and 5. However, Stochastic Gradient Descent (SGD) was chosen as it was found to converge faster than adam for this specific task. At last, a L2-regularizer was added to the loss, such that the final loss is:

$$Loss_{MSE} = \frac{1}{N_{\mathbb{1}_p}} \sum_{p=1}^{N} \mathbb{1}_p \cdot (y_p - \hat{y}_p)^2 + \lambda \frac{1}{W} \sum_{i=0}^{W} w_i^2 \tag{3}$$

with w_i the parameters of our model, W the total number of parameters and λ a hyper-parameter set to 0.001. Each epoch takes about 11 minutes and 30 seconds on the local GPU.

7.3 Results

The resulting map is post-processed to black out areas belonging to water bodies, settlements or concessions according to the land cover classification from section 6. They are now labelled as *Others*. The full regressed tile is shown on Figure 19. This is the result after 100 epochs. Visually, it does not look completely off, as one could have expected with the validation curve, but if one has a closer look, the model has several weaknesses, as it will be explained in the following paragaphs.

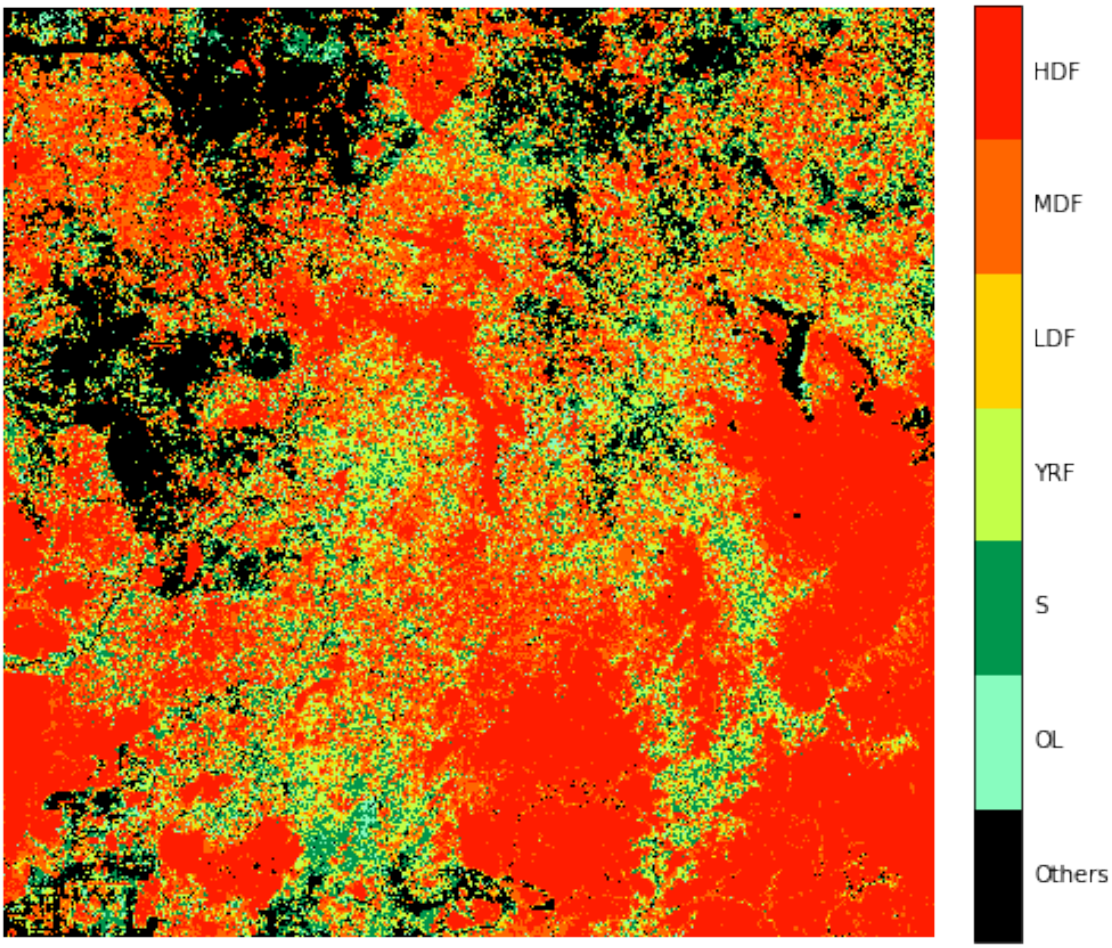


Figure 19: HCS map predicted after 100 epochs

Figure 20 is a zoom on the validation area. Here, it becomes evident that the model still suffers from inaccuracies. Open lands are particularly not well predicted by the model.

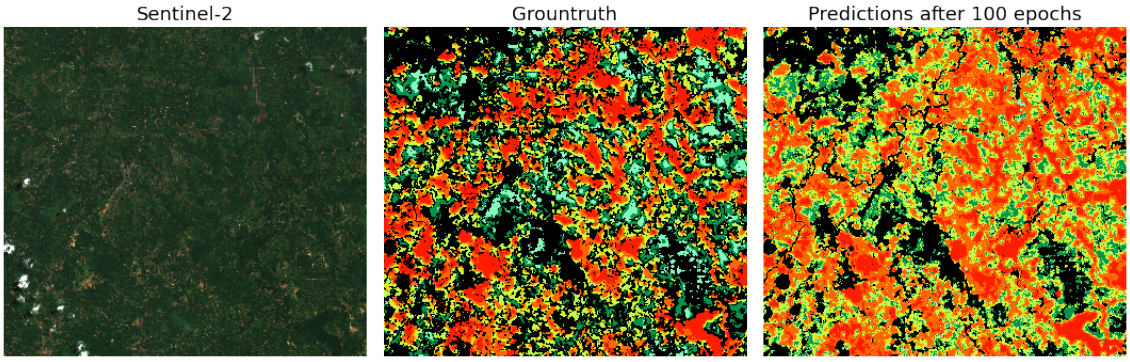


Figure 20: HCS map predicted after 100 epochs for the validation area

The prediction are evaluated with the mean average error (MAE):

$$MAE = \frac{1}{N_{\mathbb{1}_p}} \sum_{p=1}^{N} \mathbb{1}_p \cdot |y_p - \hat{y}_p| \tag{4}$$

and the root mean square error (RMSE), which is simply the square root of the MSE:

$$RMSE = \sqrt{\frac{1}{N_{\mathbb{1}_p}} \sum_{p=1}^{N} \mathbb{1}_p \cdot (y_p - \hat{y}_p)^2} \tag{5}$$

Again, the evaluation is only computed on the pixels that were either open land or vegetation according to the land cover classification. For the validation area, the MAE gives a score of 35.6 and the RMSE 45.3. As it could already be seen on Figure 20, these are not satisfactory results. Figure 21 shows the distribution of HCS values for both the groundtruth and predictions for the validation area. It is clear that there are some discrepancies between them.

The weaknesses of the model can be deduced from these distribution. First, it fails to predict low values such as the open land class. Then, the *gaussian-like* shape for values above 50 shows that it struggles to distinguish the accurate HCS values and is biased towards the overall mean. At last, the highest HCS value it predicts in the validation area is 261.5 while the highest value in the groundtruth is 341.

At last, one can plot the confusion matrix (see Figure 22) for the validation area to see how the model performs in terms of HCS classes. As the distribution plots suggested, the model can't predict the open land areas and is biased towards the mean, which is in the MDF class. The crucial aspect of the HCS Approach is the HCS threshold between

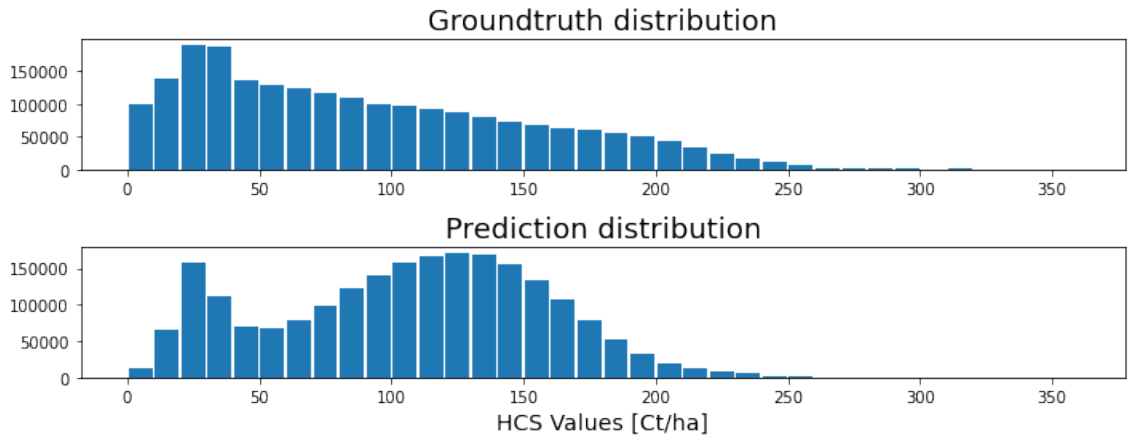


Figure 21: Distribution of HCS values in the validation area for the groundtruth and the prediction. Almost no open land (HCS<15) is predicted

scrubs and young regenerating forests. Here one can see that almost none of the forests (YRF, LDF, MDF, HDF) is predicted as open land or scrubs, which would at least mean no forest being cut due to the model's inaccuracies. On the other hand, a large part of open lands and scrubs are predicted as young regenerative forests, which would mean less land to be developed.

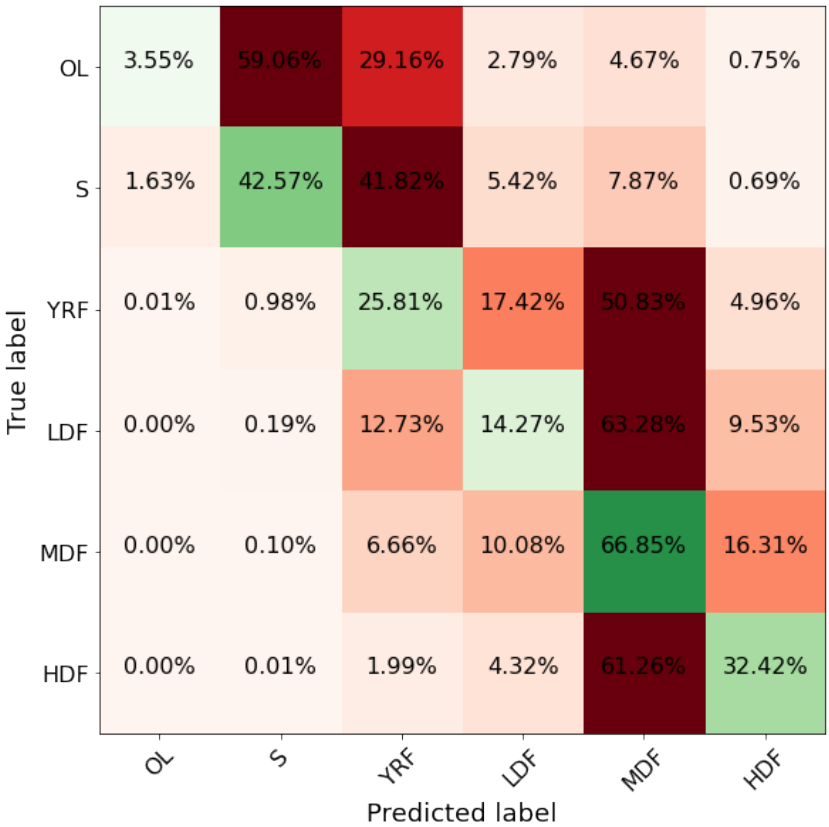


Figure 22: Confusion matrix on the validation area.

Nevertheless, it should be reminded once again that these results depend entirely on the reliability of the groundtruth, which was a combination of the land cover prediction with an interpolated and outdated biomass map. In fact, a visual inspection can sometimes bring more insights on the model performance that the statistics mentioned before. It may not be used to verify HCS values directly but it can at least be useful to check whether the HCS class predicted makes sense. For instance, Figure 23 shows a region which was neither in the training area nor in the validation one. It can be seen there that the predicted HCS classes look actually more accurate that the ESA biomass map. Figure 24 is a zoom on the smaller area indicated by the white frame in the first frame of Figure 23. Here, one can clearly see a lot of small open lands in the Sentinel-2 data which were not present in the original ESA biomass map, indicating that these forests were cut between 2017 and 2019, and confirming that the ESA biomass map is indeed outdated. On the last frame of Figure 24, one can see that they are now seen by the model and correctly predicted.

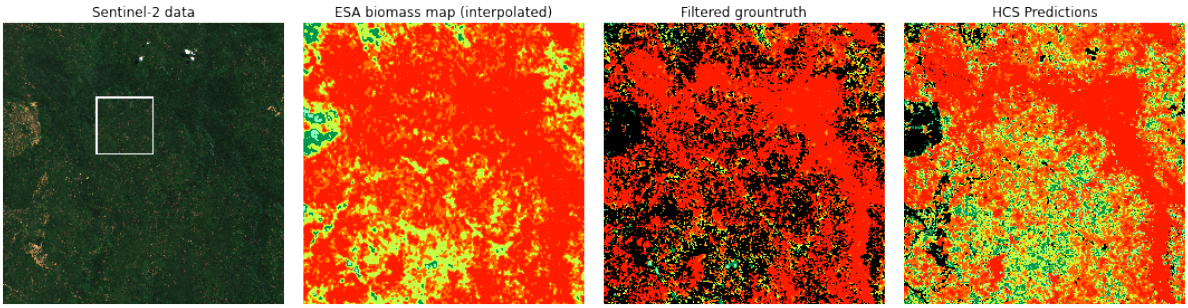


Figure 23: Patches of 30x30km^2. HCS predictions look closer to the reality than the ESA biomass map.

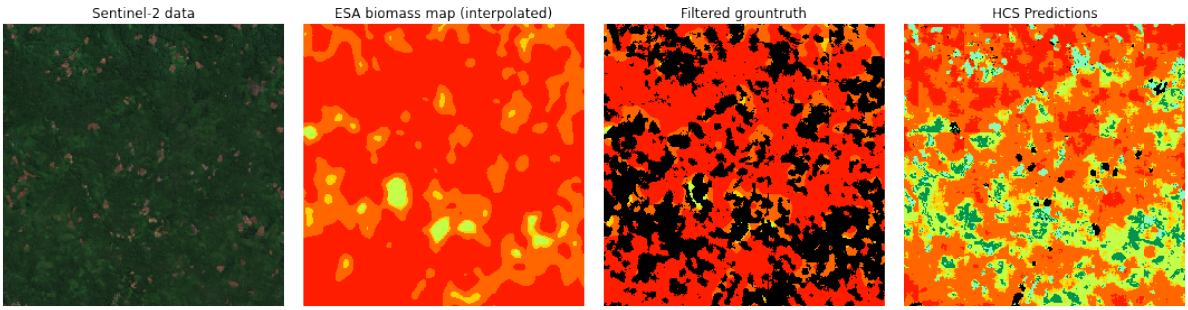


Figure 24: Patches of 6x6km^2 from white rectangle shown on Figure 23

Of course, this does not mean that the HCS predictions are accurate and one should still be very careful with these results. Nonetheless, it shows that, at least for some regions, the HCS predictions improved and updated the ESA biomass map from 2017.

8 Discussion

This last section aims at discussing the validity of the results obtained. Indeed, from the beginning, the lack of reliable groundtruth complicated greatly the project and the validation of any results obtained is still quite problematic.

For the land cover classification, the labels were made by hand and the accuracy depends thus strongly on the human performance. They were refined several times and, even is some inaccuracies persist, I believe it is now safe to assume that they are accurate enough to be considered as reliable groundtruth. Another aspect to consider is the assumption made at the beginning of the project that the labels chosen as training and validation areas summarize well enough the global land cover of the region. Indeed, the claimed accuracy of 88.23% is valid only if this assumption holds for the whole tile. Hence, if there happens to be too much diversity in this region (*e.g.* different type of vegetation or concessions, or even a whole new kind of landscape like mining for instance), one cannot know how the model will behave and this result does not hold anymore.

For the HCS regression, the issue is even bigger since our labels may actually not be reliable enough to draw conclusions uniquely by computing the accuracy or the confusion matrix. Indeed, even though it was filtered by the land cover predictions, the ESA biomass map still contains some discrepancies with the 2019 situation depicted by Sentinel-2. As mentioned at the end of last section, visual inspection is still very much needed and reveals that the HCS regression actually improves and updates the plain ESA biomass map, at least for some regions. The generated HCS map gives thus an interesting indication on how deforestation has been transforming the region for the past couple of years. Yet, the model can still be improved and these results as such cannot be validated in an accurately enough way to be considered as reliable and should consequently be kept as indicatory.

A very interesting development would be to assess the generated HCS map performance with some field measurements collected at selected locations across the region. In particular, locations close to the HCS threshold (at borders of open lands and young forests) or close to concessions would be of great interest. A few examples of such locations are shown in Appendix C.

9 Conclusion

This project aimed at estimate HCS values from Sentinel-2 images. Due to the lack of reliable groundtruth, a land cover classification was first computed to serve as a filter to improve the only HCS map at our disposal, the ESA biomass map from 2017. The land cover classification was done using Sentinel-2 images and handmade labels and an estimated accuracy of 88.23% was reached, significantly higher than the minimal accuracy required by the HCS Approach. Once the ESA map was pre-processed with the land cover prediction, a new HCS map was regressed using the same Sentinel-2 tile. The results, hardly validatable as such, seem to go towards the right direction but would need more refinement to be used for practical applications. Nonetheless, they already give good overall indications on the states of the forests and oil palm concessions in this region. An interesting development would be to validate the values predicted with field measurements. Also, only one Sentinel-2 tile was considered for this project, but it would be interesting now to expand the scope of the project and predict land cover and HCS values for the two regions initially proposed, Sanggau and Sekadau.

On a more personal note, this project gave me the opportunity to put my skills into practice and to serve a cause that I consider important. I also learnt enormously in a short amount of time. From a coding perspective, it was my first project on such a scale and also the first time I used Pytorch. I definitely learnt some important lessons like being organized in my code and keeping track of everything done. I also learnt that tiny coding errors can lead to terrible consequences and the importance of debugging quickly. Then, I discovered the fascinating world of remote sensing and satellite imagery. Coming from a micro-engineering/robotics background, all of this was very new to me and I would like now to keep exploring this field and possibly pursue other projects related that associate it with deep learning techniques.

Given the particular context of this semester, I felt really lucky that I could still work on such good conditions. Immediately after ETH shutdown, I could have access remotely to all the resources and assistance needed. Therefore, I would like to thank the EcoVision team and in particular Dr.-Ing. Jan Dirk Wegner, for giving me this opportunity, M.Sc. Nico Lang, for his advices and his earlier work on which I drew much inspiration and, last but not least, M.Sc. Nikola Kalischek, for his precious help, his constant availability and the uncountable zoom calls.

References

[1] François Chollet. "Xception: Deep Learning with Depthwise Separable Convolutions". In: *CoRR* abs/1610.02357 (2016). arXiv: 1610.02357. URL: http://arxiv.org/abs/1610.02357.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[3] Nico Lang, Konrad Schindler, and Jan Dirk Wegner. "Country-wide high-resolution vegetation height mapping with Sentinel-2". In: *Remote Sensing of Environment* 233 (Nov. 2019), p. 111347. ISSN: 0034-4257. DOI: 10.1016/j.rse.2019.111347. URL: http://dx.doi.org/10.1016/j.rse.2019.111347.

Appendices

A Handmade labels

Labels for land cover classification were made by hand. Here is the entire tile with the 1153 polygons. There were made in such a way that it is easy to cluster them into a training and a validation area. The large green polygons which are a bit outside of these clusters come from national parks and were added to ensure that some High Density forests are labelled as well.

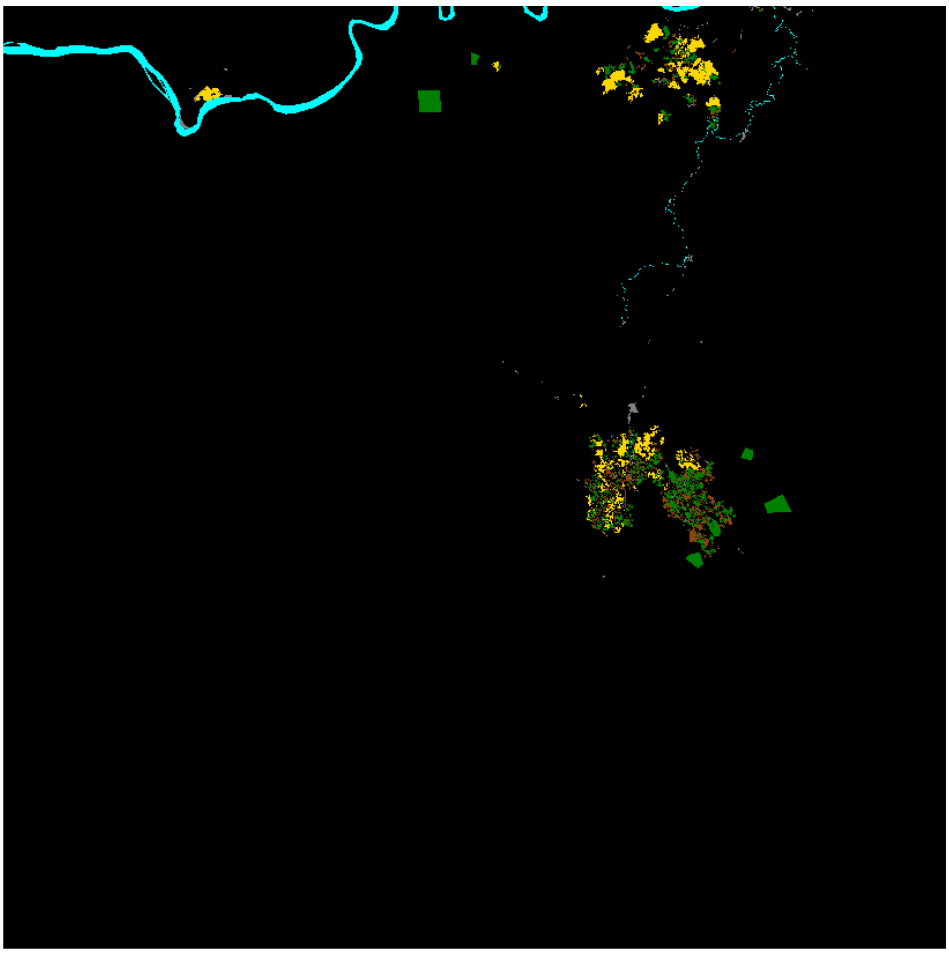


Figure 25: Hand-made labels for the full tile.

B Google Maps to help label

As Google Maps has a higher resolution, it was sometimes used to decide how to label certain areas which were not obvious visually. In particular, it turned to be very useful when differentiating native vegetation and palm trees. With the Sentinel-2 resolution, it is almost impossible to distinguish large concessions of palm trees and forests whereas one can see individual tree and their typical circular shape with the Google Maps resolution. Since the two tiles do not correspond temporally, one has to be careful using one to confirm the other. Land coverage can change very quickly in these regions. Figure 26 shows an example of the same area ($\sim$1x1km^2) seen from Sentinel-2 and Google Maps.

(a) Sentinel-2

(b) Google Maps

Figure 26: Google Maps has a higher resolution than Sentinel-2

C Examples of locations for field measurements

On-site measurements would be the ideal way to validate the results of these report. The most interesting areas are the one where the HCS threshold can be observed (*i.e.* between OL/S classes and YRF/LRF/MDF/HDF classes). Also, as HCS maps are used by companies that generally want to expand their plantations, regions near concessions are of high interest as well. Here are a few examples of locations where field measurements would be helpful to validate the HCS predictions.

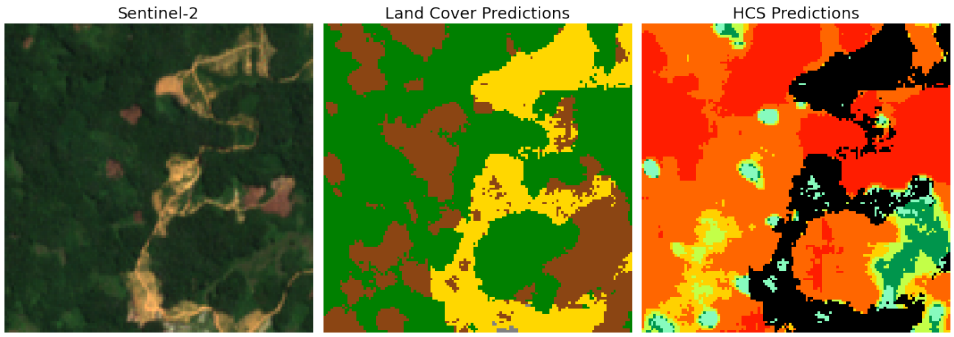


Figure 27: Patch of 1.3x1.3km^2

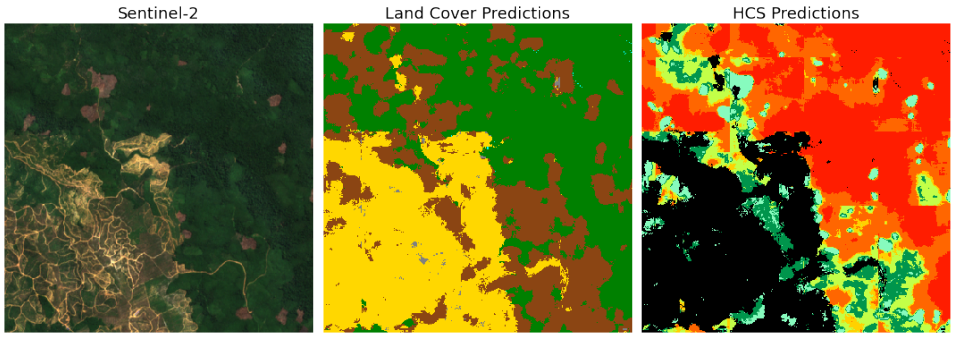


Figure 28: Patch of 4x4km^2

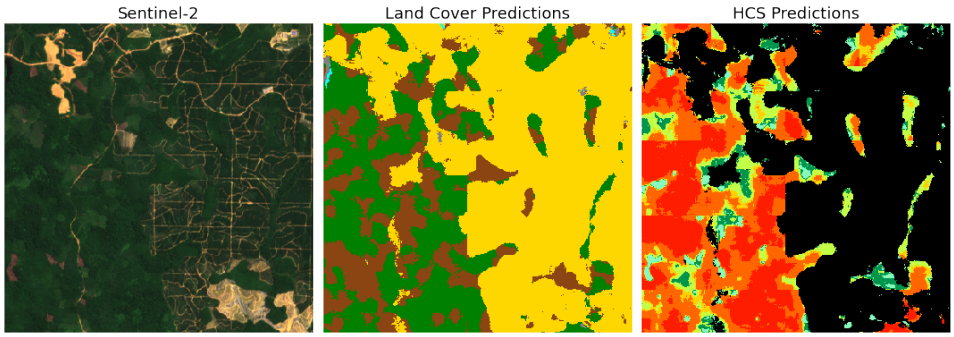


Figure 29: Patch of 4x4km^2