Can machine learning be used to generate accurate and regular Land Cover data for Switzerland from satellite imagery?

Isabel Nicholson Thomas [1]

Abstract

Understanding of environmental change in Switzerland would benefit from updated temporal resolution of land cover data. This study assesses the performance of supervised machine learning techniques to classify land cover on a regular basis in Western Switzerland using data from Landsat-5. Random Forests performs better than Multinomial Logistic Regression with an overall accuracy of 76.8% when applied to test data from a different year. However, the model requires improvement to fully capture the minority classes.

1. Introduction

Land cover is increasingly recognised as key variable for monitoring the state of the environment, providing essential data for the systematic observation of the climate and biodiversity crises in particular (Bojinski et al., 2014; Jetz et al., 2019). The availability of accurate, reliable, and timely land cover data is therefore crucial for the understanding and modelling of environmental processes (Verde et al., 2020). For example, land use and land cover change are prerequisite data for calculating of carbon emissions and storage (Li et al., 2018), feeding into international policy requirements.

Currently, the most extensive and thematically rich land cover data available for Switzerland is the ArealStatistik, derived from visual interpretation of the aerial photographs taken over a 6-year period. At present, 4 datasets are available covering the last 3 decades with a spatial resolution of 100m. These statistics are useful in terms of the detailed and country-specific categories they include, however their low update frequency and relatively coarse spatial resolution are at odds with the data needs for quantifying the dynamic and spatially variable characteristics of land cover (Ban et al., 2015). Recent efforts to downscale land cover/land use data for Switzerland have improved the spatial resolution to 25 m, however, there remains a demand for increased temporal resolution of datasets produced (Giuliani et al., 2022).

Satellite imagery provides a consistent dataset of observations which is spatially continuous and contains the temporal resolution necessary to distinguish classes with strong temporal dynamics (Verde et al., 2020). As such, the classification of remote sensing datasets has been widely accepted as the state-of-the-art in land cover mapping (Ban et al., 2015). A particular focus has been on the use of supervised machine learning methods and image classification to automate regular production of datasets with success at spatial scales from the regional to global (Inglada et al., 2017).

This study aims to assess the performance of supervised learning methods to classify land cover from satellite imagery, with the aim of presenting a method which could be applied to produce annual land cover data for Switzerland.

2. Data

2.1. Study area

The study area in the western of Switzerland is shown in Figure 1, an area of 1,570 km^2 or approximately 4% of Switzerland. This region contains the urban centers of Geneva and Lausanne, as well as part of Lake Geneva.

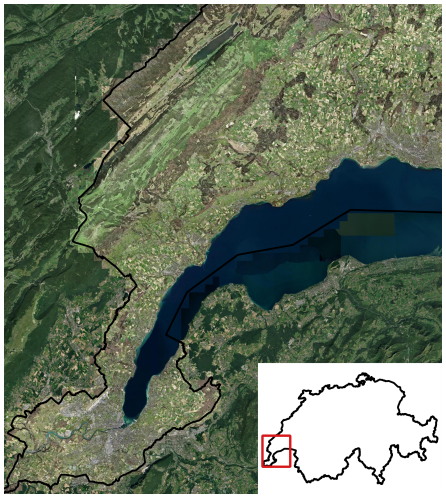

Figure 1. Study area.

[1]University of Geneva. Correspondence to: Isabel Nicholson Thomas <isabel.thomas@etu.unige.ch>.

2.2. Input data

Analysis-ready data was provided by the Swiss Data Cube which collates available satellite imagery for Switzerland (Chatenoux et al., 2021). The present study uses data from Landsat-5 which has been pre-processed to produce top-of-atmosphere reflectances. Data corresponding to the most recent land cover survey (2013-2018) comes from Landsat-8, and therefore has not been considered due to potential issues in comparison stemming from the different instrumentation used. Characteristics of the satellite data used are given in Table 1. The distribution of values of the different satellite images was assessed prior to their use to ensure extrapolation between years is feasible.

Table 1. Characteristics of satellite data.

	Resolution	
Band	Spectral	Spatial
B1 Visible	0.45 - 0.52 µm	30m
B2 Visible	0.52 - 0.60 µm	30m
B3 Visible	0.63 - 0.69 µm	30m
B4 Near-Infrared	0.76 - 0.90 µm	30m
B5 Near-Infrared	1.55 - 1.75 µm	30m

2.3. Reference data

Reference data comes from the ArealStatistik, which provides a classification for points spaced at 100m. Data from the survey periods of 1979-1985, 1992-1997, 2004-2009 was retrieved, and the corresponding satellite images for each dataset is given in Table 2. The classification used is the base level with the 6 'Principal domains' of *Artificial Areas*, *Grass & Herb Vegetation*, *Brush Vegetation*, *Tree Vegetation*, *Bare Land*, and *Watery Areas*. Reference data were reprojected into the coordinate system of the satellite imagery, and the surface reflectance values for each band extracted for each point to create the dataset.

Table 2. Reference data used for each layer of input data.

Input data	Reference data
1985/08/18 1984/07/30	1979-1985
1992/08/05 1993/07/07 1994/07/10 1996/07/31 1997/08/19	1992-1997
2004/09/07	2004-2009

3. Methods

3.1. Sampling strategy

Data from the first two survey periods (1979-1985 and 1992-1997) was used for training and testing the model, as a first iteration, with the 2004-2009 survey reserved for additional testing to assess the model's performance over time.

Random selection of training and validation data at the pixel level can lead to model evaluation over-estimating the performance of a classifier, due to the potential for spatial autocorrelation between instances (Inglada et al., 2017; Tonini et al., 2020). A potential strategy to overcome this is to select training and validation sets from separate polygons (Pelletier et al., 2016), however due to the relatively small size of the study area selected, this approach could introduce greater imbalance between LC classes. Hence, data was split using spatial k-fold cross validation, using an approach similar to the method described in (Tonini et al., 2020).

For this method, the area of interest was overlaid with a grid of 194 10km^2 cells, shown in Figure 2, and each data point was then assigned to its overlaying cell. Data for 20 cells, corresponding to around 12% of the total data, was set aside to form the test set. The cell identifier for the remaining instances which was used as the 'group' input for a grouped k-fold split, which split data into two sets for training and validation, whilst ensuring that instances for each set originated from different cells. 7 folds were used in the grouped k-fold split, to ensure consistency in the size of the validation and test sets (the size of the validation set being 1/k of the total input to a k-fold split). The final ratio was 76:12:12 for training:validation:test.

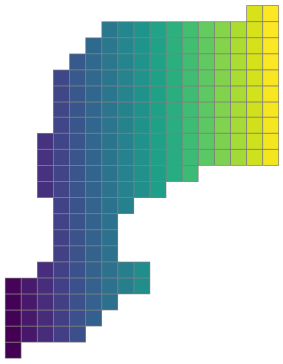

Figure 2. Spatial grid covering the study area.

The resulting class distribution of pixels used for training, validation and testing is given in Figure 3, as well as the distribution of the 2004-2009 dataset used for further testing.

3.2. Multinomial logistic regression

Multinomial logistic regression (MLR) was implemented as a simple baseline model to set a benchmark for comparison with more complex ML models.

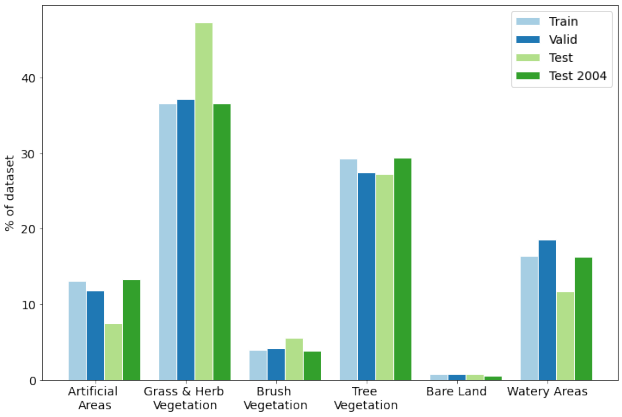

Figure 3. Class distribution across training, validation and test sets.

3.3. Random Forests

Random Forests (RF) is an ensemble learning method in which multiple decision trees are produced (Breiman, 2001). RF has become widely used in classifying remote sensing due to its high accuracy and low propensity for overfitting (Belgiu & Drăguţ, 2016) and has successfully been applied for land cover classification in similar landscapes in France using Landsat data (Pelletier et al., 2016; Inglada et al., 2017). The model was run twice, once with data as described above, and once with SMOTE oversampling. SMOTE is a data augmentation technique which synthesizes new instances of the minority classes to remove the influence of class imbalance.

3.4. Hyperparameter selection

Successive halving grid search was implemented to perform a hyperparameter search using the training and validation sets. This method begins by evaluating model performance over all hyperparameter combinations provided using a small sample of data. Under successive iterations, the best candidates from the previous round are selected and the model is evaluated using a greater number of samples.

The hyperparameters evaluated for the MLR model were the optimization algorithm used (solver) and the regularization strength (C). The range of hyperparameters used in grid search and the results of the hyperperameters leading to the best model performance, determined by highest overall accuracy, are provided in Table 3. The only optimization algorithms considered were 'SAGA' and 'SAG' due to their ability to handle large datasets.

The hyperparameters evaluated for the RF model were the number of trees in the forest (N_estimators), the maximal depth of each tree (Max_depth), the minimum samples required to split an internal node (Min_samples_split), and the minimum samples per node (Min_samples_leaf). The range of hyperparameters used in grid search and the results of the hyperperameters leading to the best model performance, determined by highest overall accuracy, are provided in Table 4. Whilst other work has included 'N_estimators' of up to 400 in hyperparameter determination, this option was omitted due to the diminishing return in increased accuracy relative to the increased computation time.

Table 3. Results of multinomial logistic regression hyperparameter search.

HYPERPARAMETER	RANGE	BEST
SOLVER	SAGA, SAG	SAG
C	100, 10, 1.0, 0.01	0.01

Table 4. Results of RF hyperparameter search.

HYPERPARAMETER	RANGE	BEST
N_ESTIMATORS	50, 100, 150, 200	200
MAX_DEPTH	10, 25, 50	25
MIN_SAMPLES_SPLIT	2, 5, 10	5
MIN_SAMPLES_LEAF	1, 10, 25, 50	10

4. Results and Discussion

Initial results in Table 5 show that the RF model performed better than the MLR model, producing an overall accuracy of 76.8% on the 2004 test data. The accuracy of both models was lower on the first set of test data taken from the first 2 time periods, likely because of this data has a notably different class distribution compared to the training and validation sets, as seen in Figure 3. Running the RF model with SMOTE did not lead to an improvement in overall accuracy (73.5% accuracy on the validation set), but rather resulted in overfitting the training data (99.2% accuracy on the training set). The overall f1 score, which represents a weighted average of the model's precision (the proportion of land cover class predictions which are correct) and recall (the proportion of on-the-ground land cover classes correctly predicted), was also calculated to give an evaluation of the model's performance considering the imbalanced classes.

Table 5. Performance metrics of the MLR and RF models.

	ACCURACY		F1 SCORE	
	MLR	RF	MLR	RF
TRAINING	65.1	79.1	60.9	77.5
VALIDATION	64.9	74.9	57.5	73.3
TEST	59.3	69.6	53.3	66.4
TEST 2004	67.6	76.8	62.5	75.2

Figure 4 shows RF model's predictions on the 2004 Landsat image test data, alongside the 2004-2009 reference data. Differences can be observed in the distribution of the vegetated

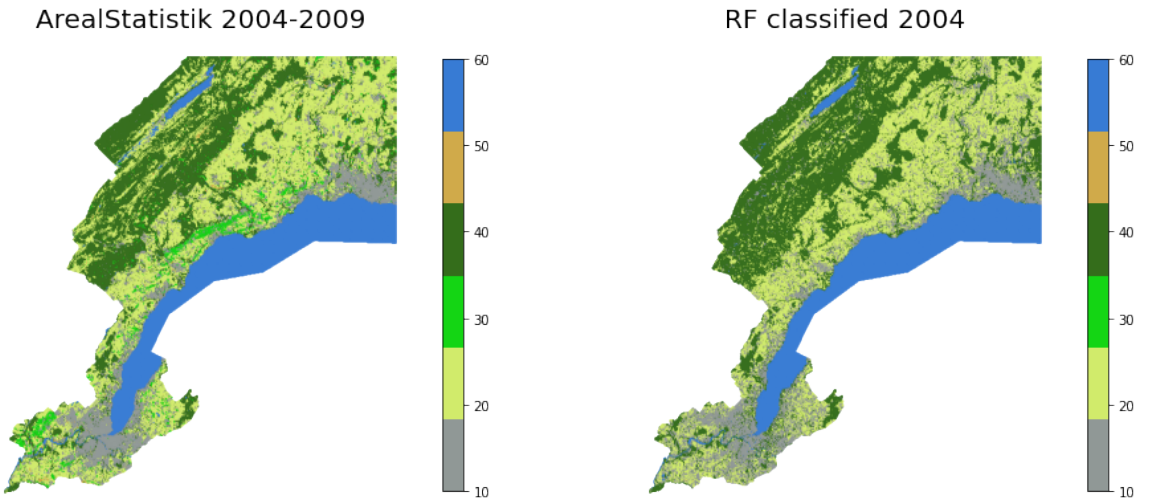

Figure 4. Reference land cover values (left) and RF predictions (right) for 2004/09/07.

land cover classes, with a distinct absence of Brush Vegetation (30) in the predictions, and an over-representation of Tree Vegetation (40) in the north-west of the study area. Slight differences in the Watery Areas (60) class to the south-west of the Lac de Joux, absent in the predictions, indicate the limitation of using reference data from a single season, as the seasonality of water presence is not well captured.

Figure 5 confirms that the model performs particularly poorly for the classification of Brush Vegetation, most frequently mis-classified as Grass & Herb Vegetation. Intra-class variability is known to be a difficulty in production of land cover maps over large areas (Inglada et al., 2017), and this is perhaps an important factor for this class, which contains diverse vegetation such as brush meadows and vines. For example, the RF model notably fails to classify the large area of Brush Vegetation (30) seen along the middle of the north-west edge of Lake Geneva, which from more detailed land cover maps can be identified as vines.

Watery areas is the best-classified class, with an f1 score of 0.97, and most likely performs well due to the distinct spectral signature of water. With higher scores for precision than recall, the presence of Artificial Areas and Tree vegetation are slightly over-estimated. Bare Land shows reasonable recall, indicating the majority actual instances of this class are captured, however the model's overall performance for this class is low due to low precision, indicating many erroneous classifications of other classes as Bare Land.

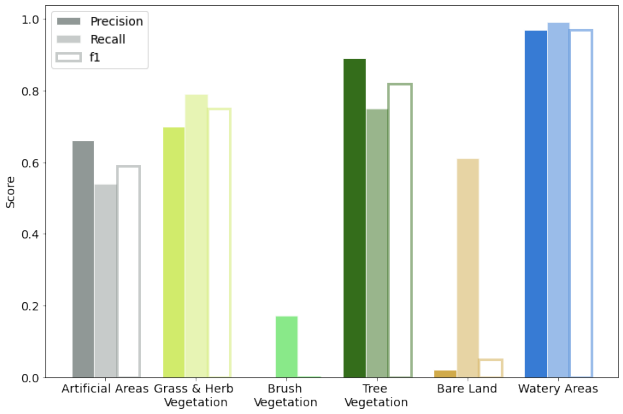

Figure 5. Performance metrics of RF classification, by land cover class.

5. Conclusions

In terms of overall accuracy, the RF model presented performs relatively well, however the poor performance on two classes raises questions on its potential applicability to other areas of Switzerland. Further work is required to determine if other methods of over- or under-sampling the dataset to deal with the class imbalance would be effective. However, additional testing with a larger study area would also be useful to increase the presence of Brush Vegetation and Bare Land in the dataset as the low number of instances may explain in part the poor performance on these classes. The classification results presented here could be improved through including spectral features (i.e. NDVI or similar indices) and temporal features (i.e. mulitple images per year) in the input data, which has previously been shown to produce a slight improvement in land cover classification accuracy. (Pelletier et al., 2016). Further testing on multiple years of the survey periods is also required to assess the impact of reference data referring to a 6-year time period.

6. Code & Data availability

The code used in this report is available at https://github.com/isabelntho/ML_final_project. Data is available at https://drive.google.com/drive/folders/1JmLc-dOOEzn11kOksIDPDhDqQotXmFBp?usp=sharing

References

Ban, Y., Gong, P., and Giri, C. Global land cover mapping using Earth observation satellite data: Recent progresses and challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:1–6, 2015. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2015.01.001. URL https://www.sciencedirect.com/science/article/pii/S0924271615000131.

Belgiu, M. and Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31, 2016. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2016.01.011. URL https://www.sciencedirect.com/science/article/pii/S0924271616000265.

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M. The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy. *Bulletin of the American Meteorological Society*, 95(9):1431 – 1443, 2014. doi: 10.1175/BAMS-D-13-00047.1. URL https://journals.ametsoc.org/view/journals/bams/95/9/bams-d-13-00047.1.xml. Place: Boston MA, USA Publisher: American Meteorological Society.

Breiman, L. Random Forests. *Machine Learning*, 45 (1):5–32, October 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL https://doi.org/10.1023/A:1010933404324.

Chatenoux, B., Richard, J.-P., Small, D., Roeoesli, C., Wingate, V., Poussin, C., Rodila, D., Peduzzi, P., Steinmeier, C., Ginzler, C., Psomas, A., Schaepman, M. E., and Giuliani, G. The Swiss data cube, analysis ready data archive using earth observations of Switzerland. *Scientific Data*, 8(1):295, November 2021. ISSN 2052-4463. doi: 10.1038/s41597-021-01076-6. URL https://doi.org/10.1038/s41597-021-01076-6.

Giuliani, G., Rodila, D., Külling, N., Maggini, R., and Lehmann, A. Downscaling Switzerland Land Use/Land Cover Data Using Nearest Neighbors and an Expert System. *Land*, 11(5), 2022. ISSN 2073-445X. doi: 10.3390/land11050615. URL https://www.mdpi.com/2073-445X/11/5/615.

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., and Rodes, I. Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sensing*, 9(1), 2017. ISSN 2072-4292. doi: 10.3390/rs9010095. URL https://www.mdpi.com/2072-4292/9/1/95.

Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., Fernandez, M., Geller, G. N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F. E., Pereira, H. M., Regan, E. C., Schmeller, D. S., and Turak, E. Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution*, 3(4):539–551, April 2019. ISSN 2397-334X. doi: 10.1038/s41559-019-0826-1. URL https://doi.org/10.1038/s41559-019-0826-1.

Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., Houghton, R. A., and Peng, S. Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015). *Earth System Science Data*, 10(1):219–234, 2018. doi: 10.5194/essd-10-219-2018. URL https://essd.copernicus.org/articles/10/219/2018/.

Pelletier, C., Valero, S., Inglada, J., Champion, N., and Dedieu, G. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187:156–168, 2016. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2016.10.010. URL https://www.sciencedirect.com/science/article/pii/S0034425716303820.

Tonini, M., D'Andrea, M., Biondi, G., Degli Esposti, S., Trucchia, A., and Fiorucci, P. A Machine Learning-Based Approach for Wildfire Susceptibility Mapping. The Case Study of the Liguria Region in Italy. *Geosciences*, 10(3), 2020. ISSN 2076-3263. doi: 10.3390/geosciences10030105. URL https://www.mdpi.com/2076-3263/10/3/105.

Verde, N., Kokkoris, I. P., Georgiadis, C., Kaimaris, D., Dimopoulos, P., Mitsopoulos, I., and Mallinis, G. National Scale Land Cover Classification for Ecosystem Services Mapping and Assessment, Using Multitemporal Copernicus EO Data and Google Earth Engine. *Remote Sensing*, 12(20), 2020. ISSN 2072-4292. doi: 10.3390/rs12203303. URL https://www.mdpi.com/2072-4292/12/20/3303.