STEPWISE REFINEMENT OF LOW RESOLUTION LABELS
FOR EARTH OBSERVATION DATA: PART 1

Daniele Cerra, Nina Merkle, Corentin Henry, Kevin Alonso, Pablo d'Angelo, Stefan Auer,
Reza Bahmanyar, Xiangtian Yuan, Ksenia Bittner, Maximilian Langheinrich, Guichen Zhang,
Miguel Pato, Jiaojiao Tian, Peter Reinartz

{FirstName.LastName}@dlr.de

Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen

ABSTRACT

This paper describes the contribution of the DLR team ranking 3rd in Track 1 of the 2020 IEEE GRSS Data Fusion Contest, with results ranking 2nd in Track 2 of the same contest are reported in a companion paper. The classifications are based on refinements of low-resolution MODIS labeling on available higher resolution Sentinel-1 and Sentinel-2 data. Results are initialized with a handcrafted decision tree integrating output from a random forest classifier, and subsequently boosted by detectors for specific classes.

1. INTRODUCTION

The 2020 Data Fusion Contest, organized by the Image Analysis and Data Fusion Technical Committee (IADF TC) of the IEEE Geoscience and Remote Sensing Society (GRSS) and the Technical University of Munich, focused on large-scale land cover mapping from globally available multimodal satellite data. The task is to train a machine learning model for global land cover mapping based on weakly annotated samples. Training, validation and test data consist of triplets of Sentinel-1 and Sentinel-2 images at 10 m/px spatial resolution, and MODIS semantic labels resampled from the original 500 m/px spatial resolution [1]. The source data are collected over different parts of the world. The 2020 contest comprised two challenge tracks: land cover classification with low-resolution labels (Track 1), and the same problem with additional high resolution semantic labels made available for training (Track 2). This paper describes the approach developed at the German Aerospace Center (DLR) ranking 3rd in Track 1. Refinements to the reported approach which resulted in a 2nd prize in Track 2 of the same contest are reported in a companion paper [2].

2. METHODOLOGY

2.1. Training Area Subset

A subset of the available training data was selected as input to the random forest classifier described in Section 2.3, using
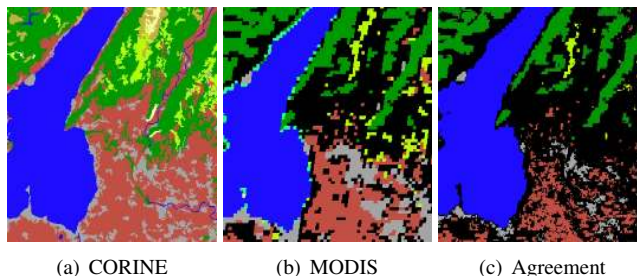
Fig. 1. Training area selection from CORINE and MODIS labels (a, b). The Random Forest classifier is trained on the intersection of the two (c). Cf. Fig. 3 for a legend.

as ground truth the agreement between MODIS labels and resampled labels from the Copernicus CORINE [3] (wherever available) or Global Land Cover Collection 2 (CGLS-LC100) [4] datasets, both at 100 m/px spatial resolution, and the SRTM Water Body Data (SWBD), at approximately 30 m/px [5]. The strongest semantic correspondences between classes in the different datasets have been researched and are reported in Table 1. An example is reported in Fig. 2.1, where the class Savanna in MODIS has been removed as it is ignored in the reference labels (appears therein in black). Higher resolution OpenStreetMap data were not used due to inconsistency with the contest's labels.

2.2. MODIS Label Refinement

The initial labeling of MODIS identifies some classes less accurately than others, and has a strong presence of the class Savanna (see Fig. 3 (b)), which should not appear in the final classification results. Exploiting the availability of higher-resolution data, the initial MODIS labeling is therefore refined by reassigning all pixels belonging to the Shrubland, Savanna, Grasslands, Wetlands, and Barren classes to one of the 8 final classes, according to a decision tree.

Initially, increasing values of Normalized Differential Vegetation Index (NDVI) in Sentinel-2 are assigned to the classes Barren, Shrubland, Grassland, Wetlands, and Forest.

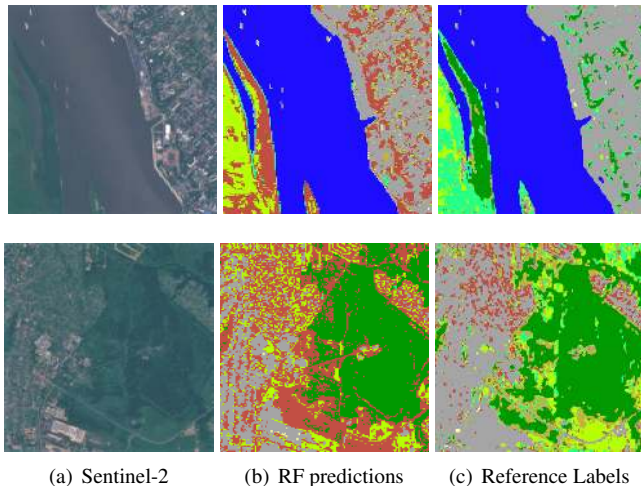**Table 1**. Semantic correspondences between classes

Class	MODIS	CGLS	CORINE
Forest	1-5	111-124	141, 311-313
Shrubland	6,7	20	322-324, 333
Savanna	8, 9	None	None
Grasslands	10	30	231, 321
Wetlands	11	90	411, 412, 421-423
Cropland	12, 14	40	211-213, 221 -223, 241-244
Urban	13	50	111, 112, 121- 124, 131-133, 142
Snow & Ice	15	70	None
Barren	16	60	331, 332, 334
Water	17	80, 81, 200	511, 512, 521-523

(a) Sentinel-2 (b) RF predictions (c) Reference Labels

Fig. 2. Random forest classifier predictions for Track 1. Cf. Fig. 3 for a legend.

The decision tree then refines the classes based on different spectral features and indices (e.g. the Normalized Water Index [6]) extracted from Sentinel-2. At this stage, pixels labeled in MODIS as Water, Cropland, Forest and Urban are mostly left unmodified. Sample results for the validation dataset are shown in column (c) of Fig. 3.

2.3. Random Forest Classification

Due to the difficulty of training neural networks with low-resolution labels and the unsatisfying results yielded by Deep Learning algorithms, Random Forest (RF) classifiers have been used instead. Optimal hyper-parameters were set using grid-search and qualitatively evaluating the results, as follows:

- Though forests of RF are generally more reliable, they did not produce satisfactory results for any class. Therefore, a single RF was used.

- Based on the implementation of the Scikit-learn python library, the following RF architecture was used: number of trees = 500, maximum depth = 100 (longest branch in a tree), maximum leaf nodes = 100. Additionally, we handled the class imbalance problem within our training data by setting the parameter "class_weight" to "balanced".

- All pixels from 22 scenes randomly drawn from the "ROIs1158_spring" training dataset were selected, using our refined MODIS labels as ground truth (see Subsection 2.1). We explain the effectiveness of the spring images due to the strong localization bias in the validation dataset, which seemed to correspond more to the spring season than to any other.

Results show confusion between the classes Cropland, Grasslands, Barren, Shrubland, Wetlands and Forest (see examples illustrated in Fig. 2). Therefore, only the predictions for the better identified classes Urban and Water were stored for the next steps. In Track 2, the predictions for classes Grasslands and Forest were better distinguished and were also employed (see companion paper describing Track 2 results).

2.4. K-means Reclassification

For the reclassification, first a global Bag of Words (BoW) is constructed. The BoW consists of patch-wise extracted sets of [MODIS label, *word*] pairs, where each *word* is a centroid obtained from a K-means classification of the union of Sentinel-1 and Sentinel-2 bands [2, 3, 4, 5, 6, 7, 8, 9, 12, 13]. The specific procedure applied to each patch is the following:

> **Data:** S1, S2, Refined MODIS
> **Result:** BoW [MODIS_Class, centroid]
> **for** *MODIS_Class in Refined MODIS* **do**
>   get **pixels** corresponding to **MODIS_Class**;
>   **if** *pixels* $\geq$ 100 **then**
>   *fit a K-means classifier using **pixels** (S1,S2);*
>   *predict **clusters** (clusters = 5);*
>   **for** *centroid in clusters* **do**
>   **if** *centroid population* $\geq$ 20% **then**
>   *include [**MODIS_Class,centroid**] in BoW;*
> **Algorithm 1:** Extraction of patch BoW

With the global BoW created, a Gaussian Naive Bayes classifier is trained and the validation set is once again classified. Sample results are shown in column (d) of Fig. 3.
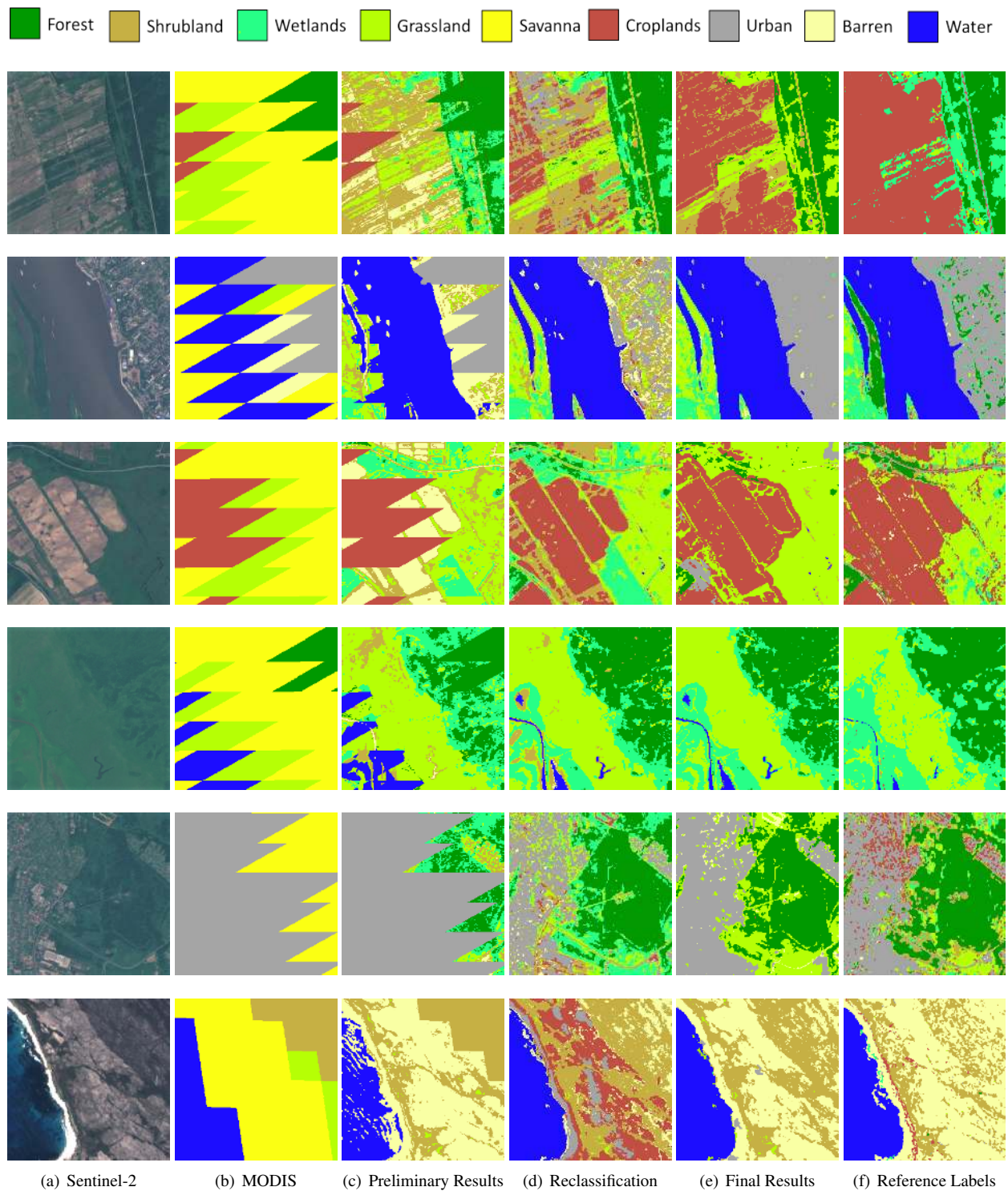
Fig. 3. Main steps of the classification strategy for Track 1. From left to right: True color combination of Sentinel-2 bands (a); coarse MODIS labelling at 500 m spatial resolution (b); output of the decision tree described in Section 2.2 (c); results after reclassification using k-means (d); final classification after post-processing (e); reference labels for selected test images (f).

2.5. Urban Detector

The Urban area as computed by the RF is refined by an urban detector using three complementary elements: Gabor texture parameters, NDTI (Normalized Difference Tillage Index) [7] computed from the Short Wave Infrared (SWIR) bands of Sentinel-2, and brightness of Sentinel-1. The Gabor features are extracted at 4 orientations and 2 scales, and capture strong texture features which are more common for Urban regions than for rural areas. Higher values of NDTI are effective at separating urban areas from bare soil. Finally, a high-intensity level in Sentinel-1 images increases the probability of having man-made structures (often partly made of metal and orthogonal surfaces). In this context, a penalty is added as additional prior if no dominant point signature (intensity above 0 dB) is present in the image patch. The resulting detection is morphologically dilated and pixels not included in the resulting area are excluded as false alarms. Residential areas, usually included in these false alarms, are recovered in a second step by checking if the NDVI of the removed pixels is above a minimum threshold, as residential areas often include vegetation in a single image element at the spatial resolution of the dataset. Finally, urban boundaries are regularized by extending the strong urban candidates to neighbouring weak candidates.

2.6. Post-processing

The Water class from the RF classifier is overlaid on the classification results. Confusion between Wetlands and Grassland is mitigated by considering local differences in NDVI and the SWIR bands. Additionally, the SWIR bands are analyzed to include neighbouring Shrubland into Croplands. Gabor texture features, already employed for the detection of Urban areas, helped in turn separating Croplands from Barren areas, as the former usually appear smoother than the latter. Morphological openings and closings have not been applied as results appeared degraded. Final results are reported in Fig. 3 (e), along with the reference high-resolution labels disclosed after the Development phase by the contest organizers in Fig. 3 (f). Results in terms of accuracy per class are reported in Table 2. Therein, Development refers to the preliminary phase assessed on the validation data, and Track 1 to the relative contest results assessed on the test data. As a comparison, the score for the best ranked teams are also reported.

3. CONCLUSIONS

This paper reports the stepwise refinement of coarse labelling for earth observation data, using as auxiliary sources multi-modal data having higher spatial resolution. The research has been conducted in the frame of the 2020 IEEE GRSS Data Fusion Contest and presents the team's classifications ranking 3rd in Track 1, where only low resolution reference labels are available. Classification results ranking 2nd in Track 2 of

Table 2. Classification results

Class	Development	Track 1
Forest	0.8534	0.8012
Shrubland	0.4621	0.3538
Grassland	0.4820	0.4446
Wetlands	0.6719	0.4273
Croplands	0.7122	0.3704
Urban	0.8605	0.8140
Barren	0.5752	0.3556
Water	0.9992	0.9838
Average Accuracy	0.7021	0.5688
Contest's Best Results	0.7073	0.5749

the contest, in which some high resolution labels are made available, are reported in a separate article [2] and build up on the results presented in this paper.

5. REFERENCES

[1] Michael Schmitt, Lloyd Haydn Hughes, Chunping Qiu, and Xiao Xiang Zhu, "Sen12ms–a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," *arXiv preprint arXiv:1906.07789*, 2019.

[2] D. Cerra, N. Merkle, C. Henry, K. Alonso, P. d'Angelo, S. Auer, R. Bahmanyar, X. Yuan, K. Bittner, M. Langheinrich, G. Zhang, M. Pato, J. Tian, and P. Reinartz, "Stepwise refinement of low resolution labels for earth observation data: Part 2," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020.

[3] M. Bossard, J. Feranec, and J. Otahel, "CORINE Land Cover Technical Guide - Addendum 2000," *European Environmental Agency, Copenhagen*, 2000.

[4] Marcel Buchhorn, Myroslava Lesiv, Nandin-Erdene Tsendbazar, Martin Herold, Luc Bertels, and Bruno Smets, "Copernicus global land cover layers—collection 2," *Remote Sensing*, vol. 12, no. 6, pp. 1044, 2020.

[5] Earth Resources Observation And Science (EROS) Center, "Srtm water body dataset," 2017.

[6] S. K. McFeeters, "The use of the normalized difference water index (NDWI) in the delineation of open water features," *International Journal of Remote Sensing*, vol. 17, no. 7, pp. 1425–1432, 1996.

[7] AP Van Deventer, AD Ward, PH Gowda, and JG Lyon, "Using thematic mapper data to identify contrasting soil plains and tillage practices," *Photogrammetric Engineering and Remote Sensing*, vol. 63, pp. 87–93, 1997.