Assessing automatically-detected changes in the post-classification change-detection of Sentinel-2 data with Visual Analytics

MIKE SIPS[1], VALERY LAVRENTIEV[1], DANIEL EGGERT[1] & JULIA NEELMEIJER[1]

Abstract: Satellite remote sensing offers the possibility to monitor the Earth's surface at high temporal and spatial resolutions. An important methodological field is the detection and interpretation of changes on the Earth's surface. A robust and widely utilized family of approaches is post-classification change-detection (PCCD). In our research, we address an important challenge to using PCCD from a user's perspective. Users often face difficulties finding changes in the result sets of PCCD that are relevant to their application scenarios. We propose a Visual Analytics approach that supports users in terms of exploring the temporal dynamics and the spatial distribution of automatically-detected changes generated via PCCD.

1 Introduction

Satellite remote sensing offers the possibility of monitoring the Earth's surface at high temporal and spatial resolutions. Mission operators increasingly publish multispectral observations of the Earth's surface in freely accessible archives, e.g., the U.S. Geological Survey (U.S. GEOLOGICAL SURVEY 2020) or the Copernicus Open Access Hub (COPERNICUS OPEN ACCESS HUB 2020). The availability of multispectral images opens opportunities for users to develop novel approaches that help human communities to address challenges such as climate change, population growth, water scarcity, and the loss of biodiversity.

An important methodological field in this context is the detection and interpretation of changes on the Earth's surface. A robust and widely utilized family of approaches for detecting changes is post-classification change-detection (PCCD) (see e.g., MAS 1992). The input for PCCD is usually a set of multispectral images describing the Earth's surface at specific points in time. PCCD then utilizes these multispectral images to construct a classified map, e.g., land-cover maps, at every time step using supervised learning methods. The utilized supervised learning methods assign a categorical label (target label), e.g. land-cover labels such as forest, urban settlement, etc. to each pixel in a multispectral image to construct these change maps. To assign target labels to pixels, these methods need some data containing ground-truth target labels. PCCD approaches assume changes at pixels where transitions in target labels are observed across two or more time steps. The output of PCCD approaches is a set of sequences representing these transitions.

[1] German Research Center for GeoSciences GFZ, Telegrafenberg, D-14473 Potsdam,
E-Mail: [sips, vlavr, daniel.eggert, neelmeijer]@gfz-potsdam.de

2 Challenge

An important challenge in using PCCD approaches is that the multispectral characteristic of a predicted target can vary from pixel to pixel, e.g., between two pixel representing a road in Berlin and Munich, or between time steps, e.g. a pixel can contain a multispectral profile of a road in winter time but can also represent a multispectral profile of a dense forest in the summer due to a dense canopy covering the road. We can attribute much of this variability in the multispectral characteristics of target labels to seasonal cycles or other long-term intra-annual changes. With the presence of the variability in multispectral characteristics, supervised learning methods often have difficulties in deriving classification models that can predict target labels precisely across space and time. Hence, the output for PCCD is usually a set of sequences representing automatically-detected changes. Users often face difficulties in finding the changes that a) describe meaningful changes, i.e. users can interpret some linear sequences as known processes at the Earth's surface, and b) are relevant for their application scenario. In our recent research project SEVA (SEVA 2020), we address this crucial challenge for PCCD. We utilize random forest classification in our Visual Analytics approach (BREIMAN 2001).

3 Visual Analytics Approach

Our Visual Analytics approach supports users in assessing sequences of transitioning target labels and deciding whether sequences are relevant for their application scenarios through interactive exploration. To decide whether sequences are relevant for their application scenarios, users need to investigate the following: a) occurrence of particular target labels within sequences in time, b) occurrence of specific transitions of target labels in sequences over time, c) distribution of target labels of particular sequences in space. To support users in their investigations of (a) - (c) through interactive exploration, we provide an overview and assessment component to users. Users can select time steps or target labels in the overview visualization for further inspection in the assessment component.

Our overview component aggregates the occurrence of target labels into a frequency distribution for each time step. We depict the corresponding frequency distribution in a bar plot (see Fig. 1(a)). The bar plot supports users in terms of understanding what target labels are involved in sequences and how frequently target labels occur at particular time steps (see (a) - occurrence of target labels in time). Thus, the bar plot supports users in selecting interesting time steps for further assessment. In addition, our approach depicts the sequence of target labels and their frequency of occurrence in an icicle plot (KRUSKAL & LANDWEHR 1983). In contrast to the original approach, we represent each time step as a circle (see Fig. 1(b)). In our icicle plot, the starting point of all sequences is the small circle in the middle of the visualization (bulls eye), and the surrounding individual rings depict the frequency of target labels in the consecutive time steps. We depict the frequency of the target labels proportional to the size available within each circle (see (b) - occurrence of specific transitions of target labels in time). Our icicle plot supports users in obtaining an overview of the major transitions and selecting potentially interesting target labels.

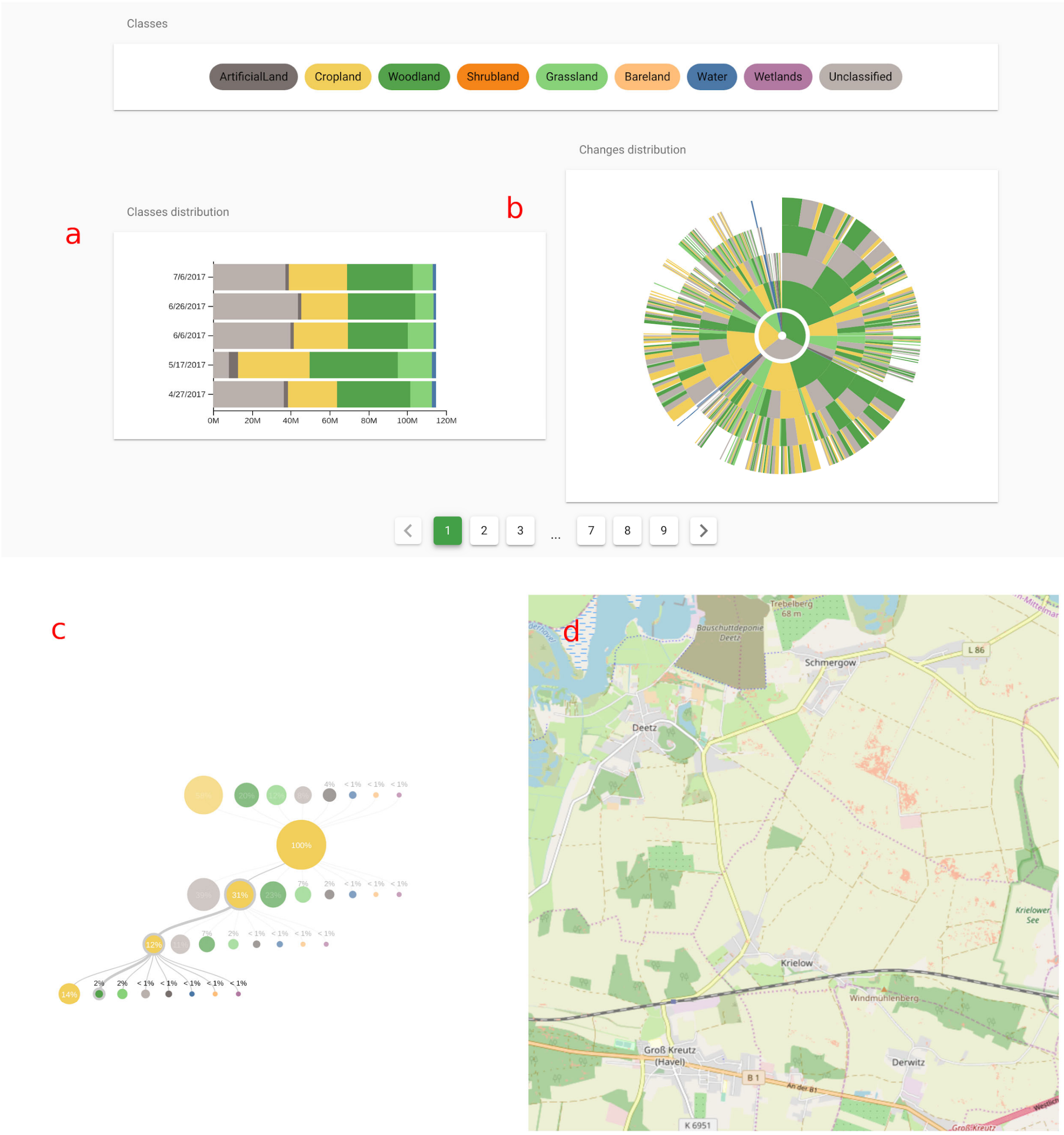

Fig. 1: Our Visual Analytics approach provides two main components for users: overview and assessment. The overview component depicts the frequency of target labels (called classes) for all time steps in a bar plot (see a) and the transition of target labels using an icicle plot (see b). The assessment component allows users to grow a sequence from a selected time step and target label into a tree visualization across time (see c). It also represents the spatial distribution of the depicted time steps and target labels of the tree in c) in the map visualization (see d).

Our assessment component depicts the selected time steps and transitions of target labels in detail. The default visualization represents the selected time step and target label as a single circle (see Fig. 1(c)). Users can explore transitions starting with the selected target label back transitions and occurrences between target labels. In addition, our assessment component depicts the spatial distributions of all target labels depicted in the tree visualization (see Fig. 1(d) and (c) - distribution of target labels of particular sequences in space).

4 Preliminary User Feedback and Future Work

The developed approach has been presented to selected experts in remote sensing for feedback regarding the usability of the tool and the usefulness for performing efficient change detection analyses. These experts confirm that our Visual Analytics approach would allow users to explore potentially interesting change sequences. They further highlight the easy-to-perform analytical interaction method, the fast response to user interaction, and the ability to explore sequences in different views in their feedback to us. In addition, the experts predicted that the ability to interactively refine individual sequences would help to reduce their daily routine work effort. In the current version, users are able to use our approach for exploring land-cover changes in Central Europe generated by PCCD on Sentinel-2 multispectral images.

The experts also provided recommendations how the Visual Analytics approach should be further extended in future projects. The three most important recommendations were: a) adding the possibility to define individual areas of interest (e.g. in form of polygons) to restrict the spatial distribution of changes, b) enabling users to assess the accuracy of the classification step in the PCCD and c) allowing users to explore changes generated by other change detection methods. These extensions would open the door to a wide range of applications in e.g., agriculture, forest and water management; landscape, habitat and urban monitoring and planning; which are all especially of interest to authorities having a variety of reporting duties on land cover changes in general.

5 References

BREIMAN, L., 2001: Random Forests. Machine Learning **45**, 5-32.

COPERNICUS OPEN ACCESS HUB, 2020: Sentinel-2 multispectral observations. https://scihub.copernicus.eu.

KRUSKAL, J. B. & LANDWEHR, J. M., 1983: Icicle Plots: Better Displays for Hierarchical Clustering. The American Statistician **37**, 162-168.

MAS, J.-F., 1999: Monitoring land-cover changes: A comparison of change detection techniques. Journal of Remote Sensing **20**, 139-152.

SEVA, 2020: Skalierbare Explorative Veränderungsanalyse für große Sentinel-2 Daten. https://www.gfz-potsdam.de/sektion/fernerkundung-und-geoinformatik/projekte/seva/

U.S. GEOLOGICAL SURVEY, 2020: Landsat multispectral observations. https://www.usgs.gov.