Analysis of intra-annual variability of time series of Sentinel 2 imagery for annual land cover mapping: A case of study in central of Portugal

William A. Martínez B.

1 Problem statement and Motivation

Mapping Land cover land use (LULC) is a fundamental task in any kind of environmental, cultural and political study, since it provides baseline for governments to undertake and monitor policies that look for sustainable livelihoods in harmony with the ecosystems. After almost 4 decades of technological evolution and powerful algorithms in mapping LULC by using earth observation, the research continues looking for adaptation of new approaches that can get benefit nowadays from big data available through new technology (i.e Sentinel 2).

The generation of annual land cover maps by using time series is a novel approach that has demonstrated superiority over methods based on static data, especially in the frame of supervised classification for the discrimination of vegetation where the intra classes variation are slightly similar (Hermosilla et al. (2015), Franklin et al. (2015). However, nowadays its application alongside Sentinel 2 technology implies different challenges. For example, an increase in the dimensionality due to the more frequent repetition of observations with different spectral frequencies over the time, thus representing a processing with large amount of images as never before. Besides that, most of the attempts for incorporating the temporal domain in the discrimination of classes have implied to make assumptions about the possible transitions of the training data. That is, generally labels belong constant over the time regardless of the natural cycles of some land cover types.

Classification, viewed from the perspective of time series, may differ from the static data approach in how the similarity between elements is computed. Generally, a time series approach takes advantage of the seasonal response of indices vegetation for different land cover types over the year to compute classification. For example, Franklin et al. (2015) shows how to classify different vegetation types by calculating metrics of normalized burn ratio, or in Vuolo et al. (2018) using simply normalized index vegetation. However, these metrics, such as the greatest disturbances year, trend magnitude year, greatest disturbance duration, among others, require the construction of dense temporal dataset for different years that at the same time disregard the rest of the spectral information available. This scenario, raises the curiosity for implementing principal components as technique to reduce dimensionality in the classification and find out possible indexes that describe not only differences in the vegetation, but also other land covers that are environmentally important such as water, soil and urban areas.

Besides dimensionality, there are also other factors that complicate the experience of land cover classification using time series. For example, no all agricultural areas for certain crop type start their sowing and harvesting at the same time (Vuolo et al., 2018), so that, the spectral profiles for the same land cover type can have different phases shift and probably also slight changes in their periods hindering the discrimination of spectral signals over the time. Moreover, when the soil moisture content provided by the rainfall in certain areas is insufficient to support the crop production the irrigation is necessary. In that case, soils with same labels can respond different to the radiance over the year due to distinct irrigation patterns.

Another important aspect that we can highlight, it is how the LULC mapping in terms of algorithms for supervised classification has disregard the capacity of the time dimension since most of the literature cover mainly random forest as classifier. For example, Vuolo et al. (2018), Franklin et al. (2015) and Hermosilla et al. (2015) show the implementation and good results that random forest classifier (RF) can bring to this problem. RF is part of the most useful and less complex mechanisms available for the classification task; important for its ability to handle high dimensionality and be insensitive to overfitting. However, the emergence of new algorithms and perspectives about how to handle temporal data structures (Bishop, 2006), raises the inquisitiveness for pairing these algorithms in the classification under the times series approach.

To continue advancing in the state of the art of classification using the time series approach alongside imagery of Sentinel 2, is therefore important to know how it is going to be explored the performance of the aforementioned proposals. In this sense, this thesis aims to use images of the Sentinel 2 from January 2017 to December 2017 and COS training dataset developed by DGT (Direção-Geral de Ordenamento do Território, Portugal). Moreover, this thesis, under the new paradigm in remote sensing of the best available pixel (BAP) composites (White et al., 2014), proposes compose images by season. This approach generally benefits classification task in areas with frequent cloudy conditions. However, in this thesis beyond that purpose, it also look for reducing the spatial variation of the crop rotation between an image and another.

2 Objectives

This thesis aims to research four questions:

- Do the integration of a index derived from principal components in time series analysis of sentinel 2 imagery in the classification task lead to a better performance than a study based on static multi-spectral data?
- Do a pixel-based image composites analysis achieve better results in classification than a scene based analysis?
- How usable is COS training data in the classification task using multi-temporal and multi-spectral data from sentinel 2 imagery?
- How do the classification approach based on multi-temporal and multi-spectral data works using random forest and support vector machine classifiers?

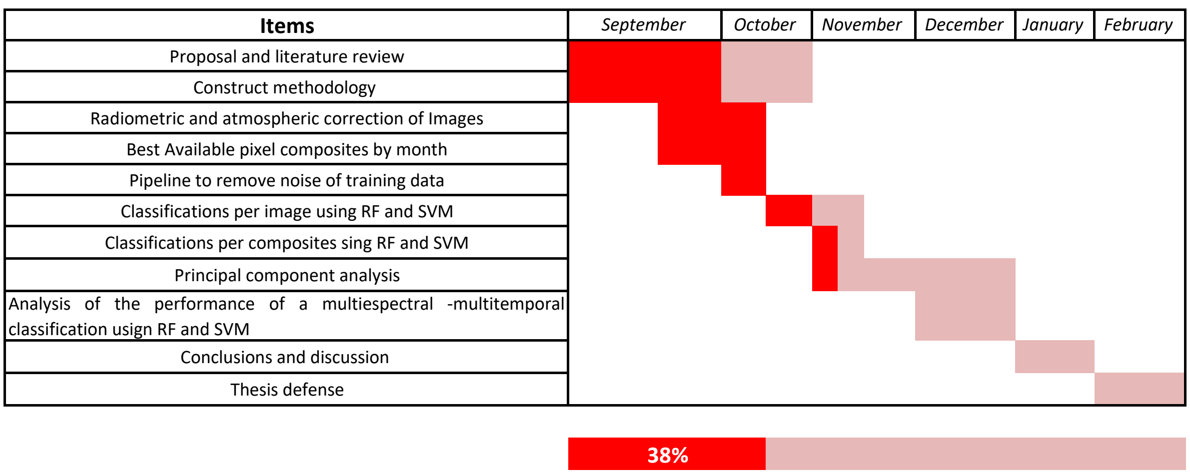

Figure 1: Cronogram

References

Bishop, C. M., 2006. Pattern recognition and machine learning (information science and statistics) springer-verlag new york. Inc. Secaucus, NJ, USA.

Franklin, S. E., Ahmed, O. S., Wulder, M. A., White, J. C., Hermosilla, T., Coops, N. C., 2015. Large area mapping of annual land cover dynamics using multitemporal change detection and classification of landsat time series data. Canadian Journal of Remote Sensing 41 (4), 293–314.

Gómez, C., White, J. C., Wulder, M. A., 2016. Optical remotely sensed time series data for land cover classification: A review. ISPRS Journal of Photogrammetry and Remote Sensing 116, 55–72.

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., 2015. An integrated landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. Remote Sensing of Environment 158, 220–234.

Jensen, J. R., Lulla, K., 1987. Introductory digital image processing: a remote sensing perspective.

Lu, D., Mausel, P., Brondizio, E., Moran, E., 2004. Change detection techniques. International journal of remote sensing 25 (12), 2365–2401.

Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C., Ng, W.-T., 2018. How much does multi-temporal sentinel-2 data improve crop type classification? International journal of applied earth observation and geoinformation 72, 122–130.

White, J., Wulder, M., Hobart, G., Luther, J., Hermosilla, T., Griffiths, P., Coops, N., Hall, R., Hostert, P., Dyk, A., et al., 2014. Pixel-based image compositing for large-area dense time series applications and science. Canadian Journal of Remote Sensing 40 (3), 192–212.