Looking at LiDAR pixel-by-pixel: a critical approach

Agnes SCHNEIDER, Philipps-Universität Marburg, Germany

Keywords: LiDAR | Archaeological Remote Sensing | machine-learning algorithms | pixel-based image analysis

The Advent of Archaeological Remote Sensing

Archaeology is known to borrow methods from other disciplines what pushes to add to its repertoire of methodology from time-to-time, opening always new and broader horizons. With the advent of remote sensing methods a decade ago, for e.g. satellite imagery, airborne digital spectral sensors, and ALS, a whole spectrum of new possibilities opened up: new ways of data acquisition and higher data quality, which on one hand pushed archaeology on the brink of (archaeological) big data and on the other hand lead to the need of new data processing and analysis methods and tools (Bennett – Cowley – De Laet 2014, Opitz – Herrmann 2018). These technological advances, the sophistication of data variety and sources, sensors and platforms improve continuously on the fly and often pose obstacles for the integration of this new kind of data into the research and thus often the potentials of these advances are not recognized and outsourced fully, due to the special requirements of the data handling. If anything these special requirements *au contraire* also push Archaeological Remote Sensing to broaden its toolkit and invoked the advent of methods borrowed from Computer Vision (like *feature detection*) which has already revolutionized for e.g. of Environmental Informatics, Climatology and Remote Sensing. Thus Archaeological Remote Sensing can outsource these disciplines in the fullest possible way and thus it is obvious that we do not have to reinvent the wheel but serve ourselves from a pretested toolset (for example as in the case of machine learning algorithms or neuronal network structures).

How to asses accessible LiDAR data?

Recently, the open access availability of LiDAR data is continuously increasing for various areas and countries. LiDAR data empowers the detection of new sites, especially in forest covered regions and the integration of archaeological sites in their landscape environment and is nowadays treated as a fundamental part of any publication or serious archaeological work. Open access LiDAR data is often acquired in different contexts and then made available to research purposes and therefore vary in coverage, quality and resolution. The (archaeologically) indigent quality and resolution of the available data poses a major problem and also the difference of data formats, like *.xyz* and *.las/laz* data. LiDAR data collected with archaeological focus can of course levels out these irregularities.

The lack of experience of using LiDAR data (apart from the technical problems) also forestalls to tap into the broad variety of LiDAR data visualizations (like Local Relief Model, Skyview factor, Multiple hillshade) which can deliver very informative delineation of the data and can already teach a great deal more about the landscape other than simple hillshade (Kokalj – Hesse 2017).

Tapping further into the possibility of LiDAR data interpretation (and of archaeological remote sensing data in general), semi-automated feature/pattern detection methods via machine learning algorithms and computer vision methods have been applied to archaeological remote sensing big data already since 2007 (Lambers – Traviglia 2016) - starting out with LiDAR data (De Boer 2007) and have shown promising results (Trier – Cowley Waldeland 2019, Verschoof-van der Vaart – Lambers 2019).

Supervised machine learning techniques (for example Support Vector Machines, Random Forests) and Deep Learning (for example (Faster) R-CNN – Convolutional Neural Networks) are currently in the focus of remote sensing archaeologists. Simultaneously two main image analysis approaches are being applied: object based image analysis (OBIA) and pixel based image analysis (Sevara – Pregesbauer – Doneus – Verhoeven – Trinks 2015). The former breaks the image down in segments, which are then classified in contrary to the latter, where pixel values (x,y,z) are used and on the basis of test-areas pixel values are assigned to certain test-classes.

Towards reproducible LiDAR data classification

This talk would like to present the preliminary results of a master's thesis project, which goal is to detect archaeological objects in LiDAR data by pixel-based image analysis. The format of the LiDAR data used is *.xyz* file format, which is how the data is received from the HVBG (Hessische Verwaltung für Bodenmanagement und Geoinformation) for archaeological use.

The study area is the vicinity of the Oppidum of the Dünsberg, because it delivers a versatile spectrum of archaeological objects, like features of Late Iron Age – Early Roman cemeteries with burial mounds and rectangular enclosures and fortified ramparts of an oppidum, specific to the Late Iron Age (of course this is only a fraction of the variability of archaeological objects). Fig. 1. displays the Local Relief Model of the study area, already very nicely emphasizing archaeological artefacts, NE to the oppidum.

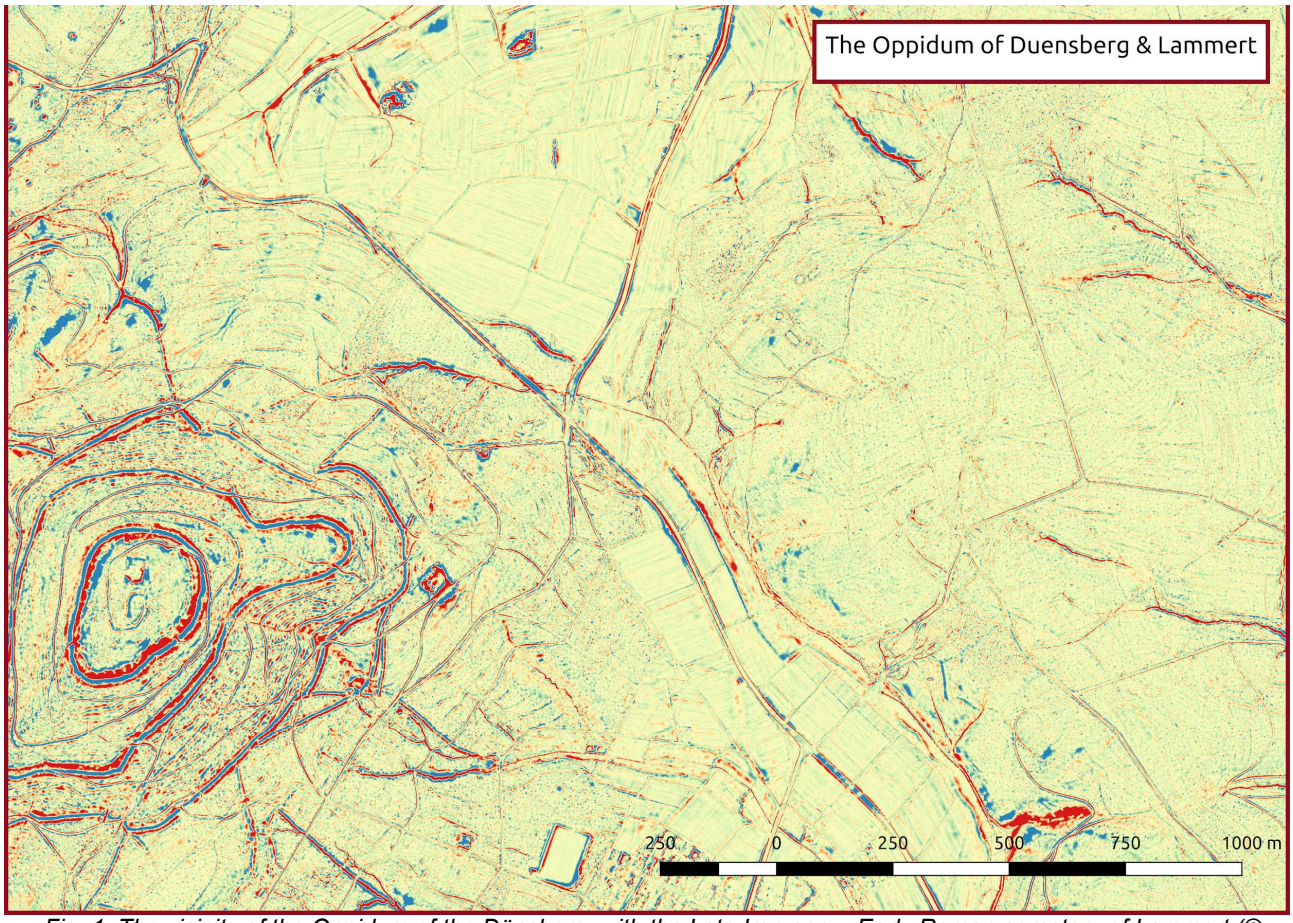

Fig. 1. The vicinity of the Oppidum of the Dünsberg, with the Late Iron age – Early Roman cemetery of Lammert (© Agnes Schneider).

In this project several machine learning algorithms are applied to the same dataset and compared to each other in R, to ensure a good comparability of the results. R is used as software environment to enable reproducibility and open access. The idea behind the project is to deliver a workflow for a smooth pixel-based classification of LiDAR data.

The first step of the workflow (Fig. 2.) consists of the preparation of the data, such as the generation of different derivatives like Local Relief model, Slope, Aspect, Negative Openness, Principal Component Analysis and Skyview Factor (I.). Following the preparation of the derivatives, a shapefile is generated containing the training areas, assigned to certain classes . The derivatives (also called predictors) and the (rasterized) shapefile are stacked together to enable the extraction of the pixel values "along" the shapefile

for the training areas, that is for the different classes into a valuable containing containing pixel values for the defined classes (II./III.).

The subsequent step (III.) applies the machine learning algorithm which creates a model on the basis of the training areas. Then on the basis of the model the prediction is being made.

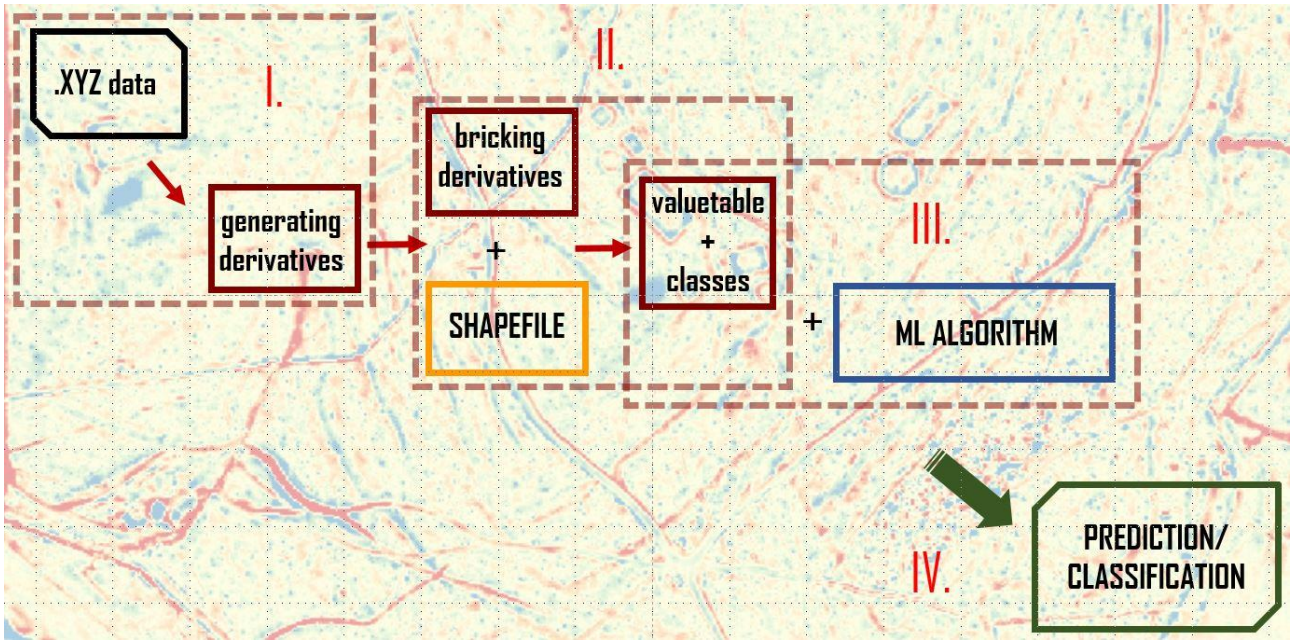

Fig. 2. The workflow of the project (© Agnes Schneider).

As already mentioned, several (linear and nonlinear) machine learning algorithms such as Support Vector Machines, Classification and Regression Trees and Random Forest for eg. Are compared but also shapefiles with different number of classes are compared and analysed to understand how to achieve the best results.

Along the way it became clear that LiDAR data with 1 point per m2 does not have a very high resolution, but on the other hand (after calculating DTMs) it also leaves us with artifacts in the data (visible in Fig 1. & 2.).Thus the data quality has to be dealt with, otherwise a pixel could belong to more classes at the same time. At the moment, ways to improve the data basis are explored to refine the prediction of archaeological objects. Connected to this, conclusively the viability of pixel based image analysis of LiDAR data is going to be addressed and assessed in this talk.

References

Bennett R. – Cowley, D – De Laet, V. (2014) The data explosion: tackling the taboo of automatic feature recognition in airborne survey data. ANTIQUITY 88 (2014): 896–905.

De Boer A. 2007. Using pattern recognition to search LIDAR data for archeological sites. In: The World is in Your Eyes. CAA2005. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 33rd Conference, Tomar, March 2005, Figueiredo A, Leite Velho G (eds.). Tomar: CAA Portugal; 245–254.

Kokalj – Hesse 2017. Airborne Laser Scanning Raster data Visualization. A Guide to Good Practice. Ljubljana 2017.

Lambers – Traviglia 2016. Automated detection in remote sensing archaeology: a reading list. AARGnews 53,2 (2016) 25-29.

Opitz, R. and Herrmann, J. (2018). Recent Trends and Long-standing Problems in Archaeological Remote Sensing. Journal of Computer Applications in Archaeology, 1(1), pp. 19–41, DOI: https://doi.org/10.5334/jcaa.11

Sevara – Pregesbauer – Doneus – Verhoeven – Trinks 2015. Pixel versus object – A comparison of strategies for the semi-automated mapping of archaeological features using airborne laser scanning data. Journal of Archaeological Science: Reports 5 (2016) 485–498. http://dx.doi.org/10.1016/j.jasrep.2015.12.023

Trier – Cowley – Waldeland 2019. Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland. Archaeological Prospection 2019, 165-175. https://doi.org/10.1002/arp.1731

Verschoof-van der Vaart, W.B. and Lambers, K., 2019. Learning to Look at LiDAR: The Use of R-CNN in the Automated Detection of Archaeological Objects in LiDAR Data from the Netherlands. *Journal of Computer Applications in Archaeology*, 2(1), pp.31–40. DOI: http://doi.org/10.5334/jcaa.32