O03: Earth Observation Data Analysis

Lectures

1. Introduction into the Case Study https://www.youtube.com/watch?v=u9n3hkuwO2g
2. Predictive Modelling https://www.youtube.com/watch?v=X1H8g_LaNQ0

Lecture Questions

As machine learning and spatial modelling is probably a new field for you, try to make sense of all the important terms first:

- What are predictors and what is a response variable?
- What problems can occur if we say a model is “overfitted”?
- How can one detect if a model is overfitted?
- What is the difference between a classification and a regression model?

Reading Task

Read the Article "Decision-Tree, Rule-Based, and Random Forest Classification of High-Resolution Multispectral Imagery for Wetland Mapping and Inventory" by Tedros M. Berhane et al. (2018). You can download the article here as it is published as open access.

While the topic of this article is about wetland mapping (i.e. called a categorial map in the article yesterday), it is also a very nice introduction to the machine learning workflow in a practical example without being too technical. If you do not understand the text word for word, do not worry - we will discuss the main outcomes and messages in person next week. For now focus on the main ideas and try to answer the following questions:

- What is a decision tree?
- What is the difference between a decision tree and a rule-based classification?
- What was the field data used for in this study?
- How was the remote sensing data processed prior to the classification?
- What positive and negative aspects do the authors identify for the usage of a random forest model for classifications?

R Practice

To put the modelling concepts into practice, please download these resources and find the Rscript `TASK-modelling.R` and data `fogo_plots.csv`. Open the .R file in Rstudio. IIt will guide you through to calculate a linear model. If this was easy for you, you can continue with the R practice in O04: Current Applications of Remote Sensing where you will learn to train a random forest model.

Lecture References

Baker, M. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454 (2016). https://doi.org/10.1038/533452a

Blanco, C. M. G.; Gomez, V. M. B.; Crespo, P. & Ließ, M.Spatial prediction of soil water retention in a Páramo landscape: Methodological insight into machine learning using random forest Geoderma, Elsevier BV, 2018, 316, 100-114

Börner K., Boyack K.W., Milojević S., Morris S. (2012) An Introduction to Modeling Science: Basic Model Types, Key Definitions, and a General Framework for the Comparison of Process Models. In: Scharnhorst A., Börner K., van den Besselaar P. (eds) Models of Science Dynamics. Understanding Complex Systems. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23068-4_1

Breiman, L.Random Forests Machine Learning, Springer Science and Business Media LLC, 2001, 45, 5-32

Hansen, M. C.; Potapov, P. V.; Moore, R.; Hancher, M.; Turubanova, S. A.; Tyukavina, A.; Thau, D.; Stehman, S. V.; Goetz, S. J.; Loveland, T. R.; Kommareddy, A.; Egorov, A.; Chini, L.; Justice, C. O. & Townshend, J. R. G.High-Resolution Global Maps of 21st-Century Forest Cover Change Science, American Association for the Advancement of Science (AAAS), 2013, 342, 850-853

Maxwell, A. E.; Warner, T. A. & Fang, F.Implementation of machine-learning classification in remote sensing: an applied review International Journal of Remote Sensing, Informa UK Limited, 2018, 39, 2784-2817

Meyer, H.; Katurji, M.; Appelhans, T.; Müller, M.; Nauss, T.; Roudier, P. & Zawar-Reza, P.Mapping Daily Air Temperature for Antarctica Based on MODIS LST Remote Sensing, MDPI AG, 2016, 8, 732

Yaseen, Z. M.; Al-Juboori, A. M.; Beyaztas, U.; Al-Ansari, N.; Chau, K.-W.; Qi, C.; Ali, M.; Salih, S. Q. & Shahid, S.Prediction of evaporation in arid and semi-arid regions: a comparative study using different machine learning models Engineering Applications of Computational Fluid Mechanics, Informa UK Limited, 2019, 14, 70-89