



# Comparison of machine learning algorithms for fuzzy land cover classification

Dainius Masiliūnas, Nandin-Erdene Tsendbazar, Jan Verbesselt

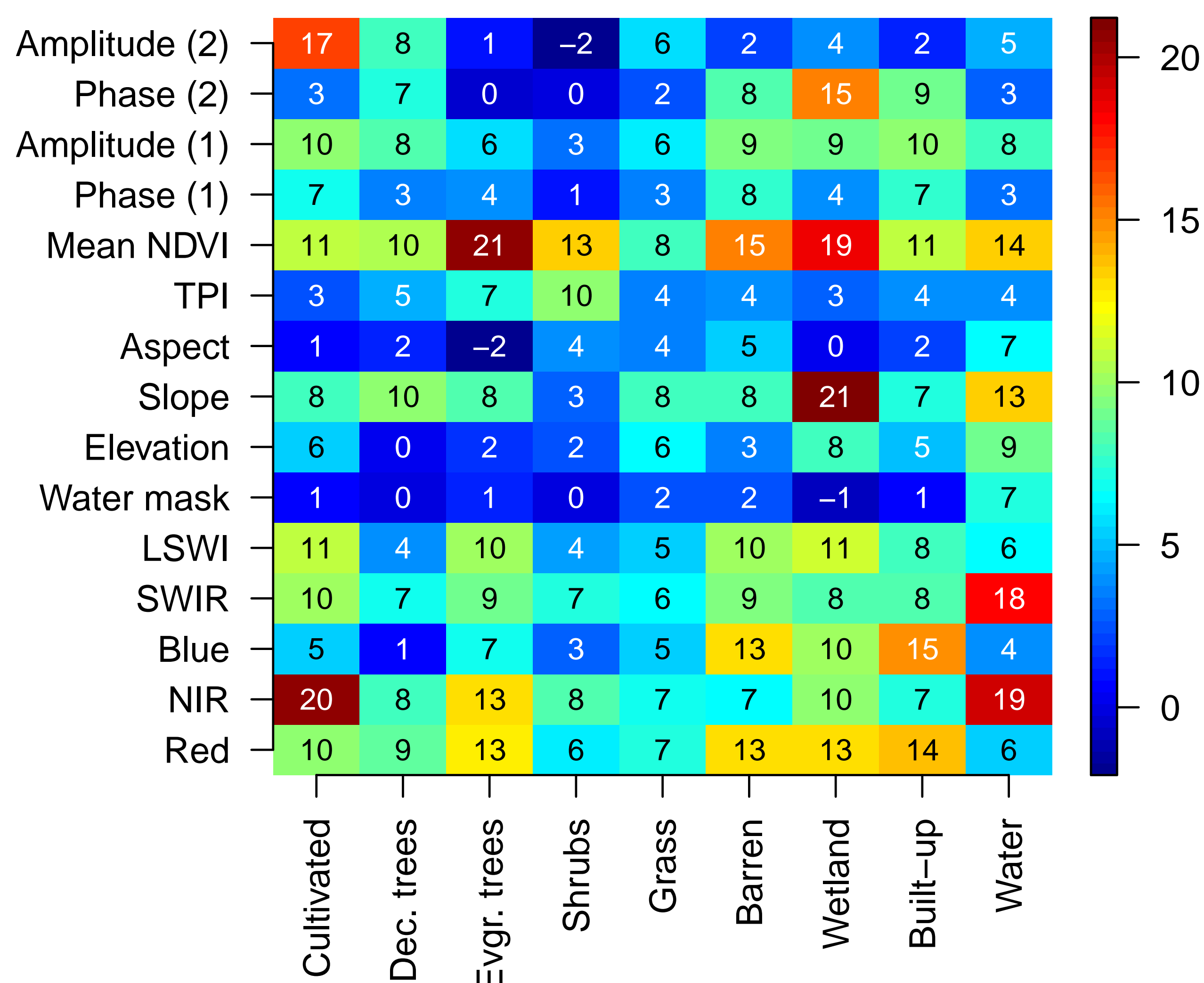
## 1. Abstract

The current standard of land cover classification is to assign each pixel to one land cover class, which at coarse resolution causes loss of information about mixed land cover. Fuzzy land cover classification, which assigns fractions of each land cover class to each pixel, can deal with mixed pixels. However, so far its application has been limited to city-scale areas and four or fewer classes.

In this paper, the classification accuracy and processing speed were compared for three fuzzy classification machine learning algorithms: random forest regression, fuzzy c-means, neural networks. The algorithms were used to classify the whole boreal-temperate forest gradient zone between Finland and Lithuania into nine land cover classes. Results showed that all of the tested algorithms are able to achieve similarly high classification accuracy. Random forest regression accuracy was the highest, but its processing speed was the lowest. The results are a milestone for creating a global fuzzy land cover classification product incorporating user-specific requirements.

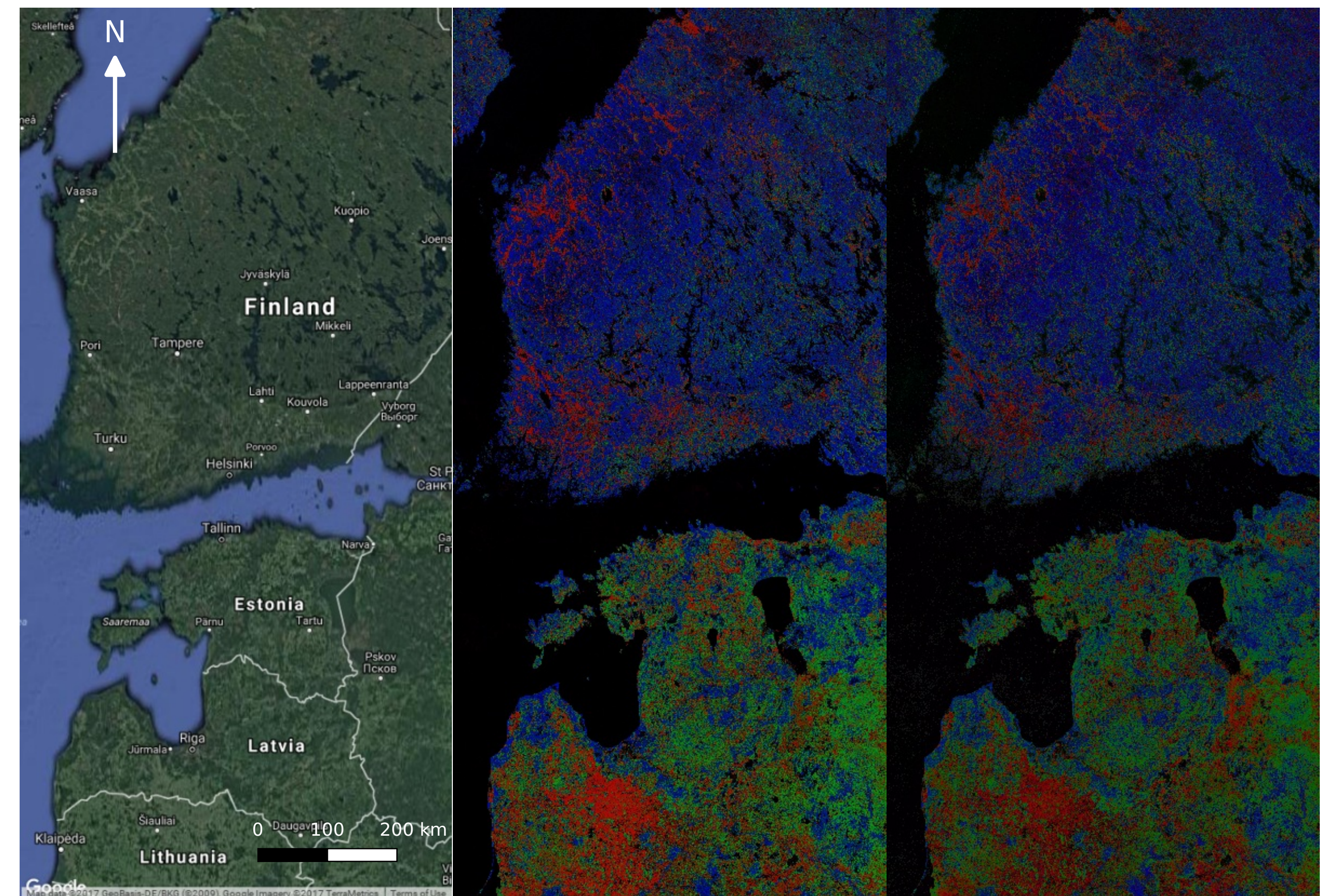
## 2. Data and methods

6 spectral and 4 temporal metrics were derived from the full 100 m Proba-V surface reflectance archive over tile X20Y01. EDEM was used to derive 4 terrestrial metrics. In addition, a water mask was derived from the Proba-V quality control layers. These 15 metrics (see figure 1, Y axis) were input into machine learning models (fuzzy c-means, random forest regression, neural networks) for predicting the fraction of each of the 9 land cover classes in each pixel: cultivated land, deciduous trees, evergreen trees, shrubs, grasslands, barren, wetland, built-up and water. The models were trained and validated on 480 sample points collected through manual high-resolution satellite and aerial image interpretation. All processing was done on a Proba-V MEP virtual machine with 32 threads and 32 GiB RAM.

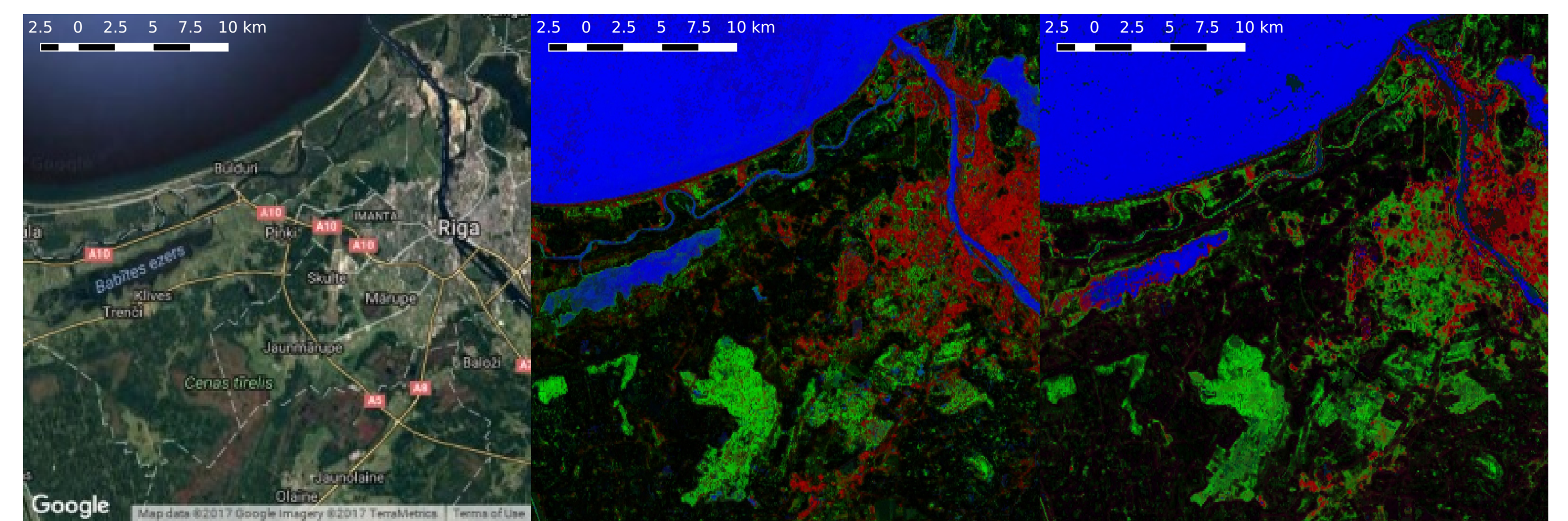


**Figure 1.** Covariate permutation importance for holdout random forest regression. X axis: all covariates used in this study, Y axis: all classes used in this study. Values indicate the increase in prediction RMSE when a given covariate is shuffled: higher values mean higher importance of the covariate.

## 3. Results



**Figure 2.** Results of full-tile fuzzy classification. Left: true colour Google imagery of the study area; middle: random forest regression algorithm; right: neural network algorithm. The colours represent classes: cultivated land (red), deciduous trees (green), evergreen trees (blue).



**Figure 3.** Close-up of fuzzy classification in the area surrounding the city of Riga, Latvia. Left: true colour Google imagery; middle: random forest regression algorithm; right: neural network algorithm. The colours represent classes: built-up (red), wetlands (green), water (blue).

## 4. Conclusion

Random forest regression was the most accurate (MAE: 11.7%), but not significantly more than fuzzy c-means (MAE: 12.0%) and neural networks (MAE: 12.7%). The prediction times were very different: 1229, 42 and 25 minutes respectively to process an entire tile. Thus random forest regression was the most accurate but slowest algorithm, neural networks was the least accurate and fastest. All of these algorithms produced spatially consistent output. All types of covariates were important to include in the models, with the exception of terrain aspect and the Proba-V water mask (see figure 1).

## 5. Discussion

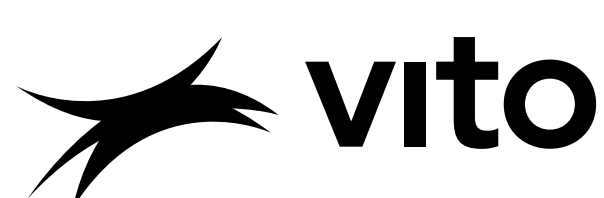
Fractional land cover mapping is capable of representing smooth transitions between classes in space (and time), and users can produce customised discrete maps by defining their own rules. Larger scale application of fuzzy classification is needed to test whether it is computationally feasible for global mapping and whether model accuracies change depending on the area of interest. This will be investigated in the future as part of the Copernicus Global Land Operations "Vegetation and Energy" (CGLOPS) project.

## 6. Acknowledgements

We would like to thank the MEP-TPS project and VITO for providing support for and access to data and computational resources on the Proba-V Mission Exploitation Platform (MEP).



Dainius Masiliūnas  
Email: [dainius.masiliunas@wur.nl](mailto:dainius.masiliunas@wur.nl)  
LinkedIn: <https://www.linkedin.com/in/greatemerald>  
ResearchGate: [https://www.researchgate.net/profile/Dainius\\_Masiliunas](https://www.researchgate.net/profile/Dainius_Masiliunas)  
Laboratory of Geo-information Science and Remote Sensing, <http://www.wur.nl/grs>  
P.O. Box 47, 6700 AA Wageningen



Read full text online:

<https://edepot.wur.nl/424175>

Source code:

<https://github.com/GreatEmerald/master-classification>

