Pareto-optimal model and threshold selection approach for biased support vector machine in tropical tree species mapping

Rami Piiroinen [1,2]*, **Fabian Ewald Fassnacht** [3], **Eduardo Maeda** [4], **Janne Heiskanen** [2,1], **Benjamin Mack** [5], **Jouko Rikkinen** [6,7,8] **and Petri Pellikka** [1,2,8]

[1] Earth Change Observation Laboratory, Department of Geosciences and Geography, University of Helsinki, P.O. Box 64, FI-00014, Helsinki, Finland; janne.heiskanen@helsinki.fi (J.H.); petri.pellikka@helsinki.fi (P.P.)

[2] Institute for Atmospheric and Earth System Research, University of Helsinki

[3] Institute of Geography and Geoecology, Karlsruhe Institute of Technology, Kaiserstraße 12, 76131 Karlsruhe, Germany; fabian.fassnacht@kit.edu (F.M.)

[4] Fisheries and Environmental Management Group, Department of Environmental Sciences, University of Helsinki, P.O. Box 68, FI-00014, Helsinki, Finland; eduardo.maeda@helsinki.fi (E.M.)

[5] Institute of Geographical Sciences – Remote Sensing and Geoinformatics, Freie Universität Berlin, Berlin, Germany.

[6] Faculty of Biological and Environmental Sciences, University of Helsinki

[7] Finnish Museum of Natural History Luomus, University of Helsinki

[8] Helsinki Institute of Sustainability Science, University of Helsinki

* Correspondence: rami.piiroinen@helsinki.fi; Tel.: +358-45-632-5313 (R.P.)

Eucalyptus spp. and *Acacia mearnsii* are common exotic tree species in eastern Africa. *Acacia mearnsii* is considered a highly invasive species that is replacing native species and *Eucalyptus* spp. is known to consume high amounts of ground water with suspected effects on native flora. Mapping the occurrence of these species in the Taita Hills, Kenya (part of Eastern Arc Mountains Biodiversity Hotspot) is important as there is lack of knowledge on their occurrence and ecological impact in the area. Mapping methods that require a lot of fieldwork are impractical in areas like the Taita Hills, where the terrain is rugged and the infrastructure is poor. Our aim was hence to map the occurrence of these tree species in a 100 km² area using airborne imaging spectroscopy and laser scanning. We used one class biased support vector machine classifier (BSVM) as it needs labeled training data only for the positive classes (*A. mearnsii* and *Eucalyptus* spp.), which potentially reduces the amount of required field work. In this study the field data for all tree species was available for validating the results. We introduce a new approach for parameterizing and setting the threshold level simultaneously for the BSVM classifier. The introduced approach outperformed other commonly used techniques. The crown level F1-score was 75.2 for *Eucalyptus* spp. and 76.8 for *A. mearnsii*. Both species are particularly located on steeper slopes and higher altitudes and have notable occurrence near the remaining native forest fragments.