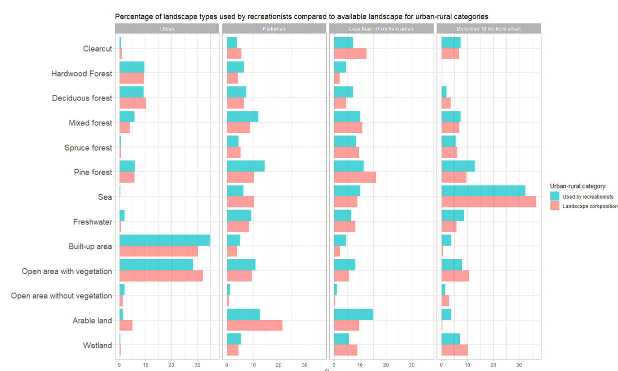
95 In search of a human habitat: using machine learning to explore the role of landscape characteristics in human outdoor recreation

Carl Lehto, **Marcus Hedblom**, **Erik Öckinger**, **Thomas Ranius**, Swedish University of Agricultural Sciences, Sweden

As the importance of outdoor recreation increasingly has been recognized due to its positive effect on human well-being and health there has been a renewed focus on how to ensure that the natural and cultural landscape can produce sufficient recreational opportunities. This is especially true in urban environments, where high land use pressure due to urbanisation often has lead to the loss of green space. To ensure that the managed landscape can supply recreational opportunities requires an understanding of what landscape characteristics (such as type and composition of land cover, topology and heterogeneity) are drivers of different kinds of outdoor recreation. Previous research in the field has to a large degree focused on establishing preferences of different kinds of environments e.g. by showing people photographs and asking questions (Gundersen and Frivold 2008); recently an increasing number of studies have been employing Public Participatory GIS-approaches to collect large amounts of data on human landscape usage (e.g Korpilo, Virtanen, and Lehvävirta (2017)). Still, most such studies are linked to specific areas (e.g. a single national park) or only looking at specific features (e.g. forest type, openness, heterogeneity).

In this study, we embraced the PPGIS approach on a large scale in order to explore the question of what landscape factors are most important for an area to be chosen for recreation. Employing a digital survey sent to a representative sample of residents of Sweden, 2856 respondents pinpointed the location of their latest outdoors recreational visit on a map and also provided details of the visit (such as type of activity, the time spent on location, distance travelled from home, etc.). Demographic information on the respondents were also collected. The data was initially analyzed in an exploratory manner, looking e.g. at the travel distances for different types of activites and which land cover types are selected for along the gradient between urban and rural areas using the Manly-chesson selection index.

The main analysis utilised machine learning in a used/available framework, where for each location used by a respondent a random location within that respondents travel distance was picked as a sample of what environment that person had available to them. For each location (both used and available) a number of map covariates were extracted from the surrounding area, e.g. land cover type, topology and presence of paths and roads. These map covariates were then used as predictors along with the demographic variables and the data on the recreational activities. Some variables were also combined to create new predictor variables, such as using reclassified land cover data to estimate landscape heterogeneity through the use of the Q index (Díaz-Varela, Roces-Díaz, and Álvarez-Álvarez 2016). Five different models were created looking at different scales of landscape.



Modelling was performed with boosted regression trees (BRT), a machine learning method of the gradient boosting class (Elith, Leathwick, and Hastie 2008). BRT have been shown to create models with high degrees of predictive power, and can handle any number of predictors and interactions between predictors. It is a powerful tool to explore large datasets, and especially useful in that you do not need to specify interactions a priori, nor is there a need for model selection processes. A weakness of the method is that the models can be harder to interpret than traditional regression models;

resulting in ‘black boxes’ that are very good at predicting to new data but hard to understand. However, recent advances in the field have yielded methods to increase the interpretability of these type of models (Molnar 2018).

To our surprise, all models performed poorly at distinguishing the used sample from the availability sample, with cross-validated AUC values between 0.54-0.567, meaning the models performed only slightly better than chance. This suggests that land cover type and composition, topology or other spatial factors were not influential in the choice of recreational area. Neither were any patterns found linked to demography (e.g. gender, age, education or living in urban or rural areas), implying that preferences are rather homogenous across the surveyed population. We argue that these results should not be interpreted to mean that the characteristics of the landscape does not matter for outdoor recreation, but instead that other factors (that were not included in our models) could be more important.

References

Díaz-Varela et al. 2016**. https://doi.org/10.1016/j.landurbplan.2016.05.004.** Elith et al. 2008**. https://doi.org/10.1111/j.1365-2656.2008.01390.x.** Gundersen VS & Frivold LH. 2008**. https://doi.org/10.1016/j.ufug.2008.05.001.** Korpilo et al. 2017**. https://doi.org/10.1016/j.landurbplan.2016.08.005.** Molnar, C. 2018**. https://doi.org/10.21105/joss.00786.**