From data to decisions: Improving land use models with machine learning for sustainable land management

Simon Thomsen & Melvin Lippe

Background and Motivation of Study:

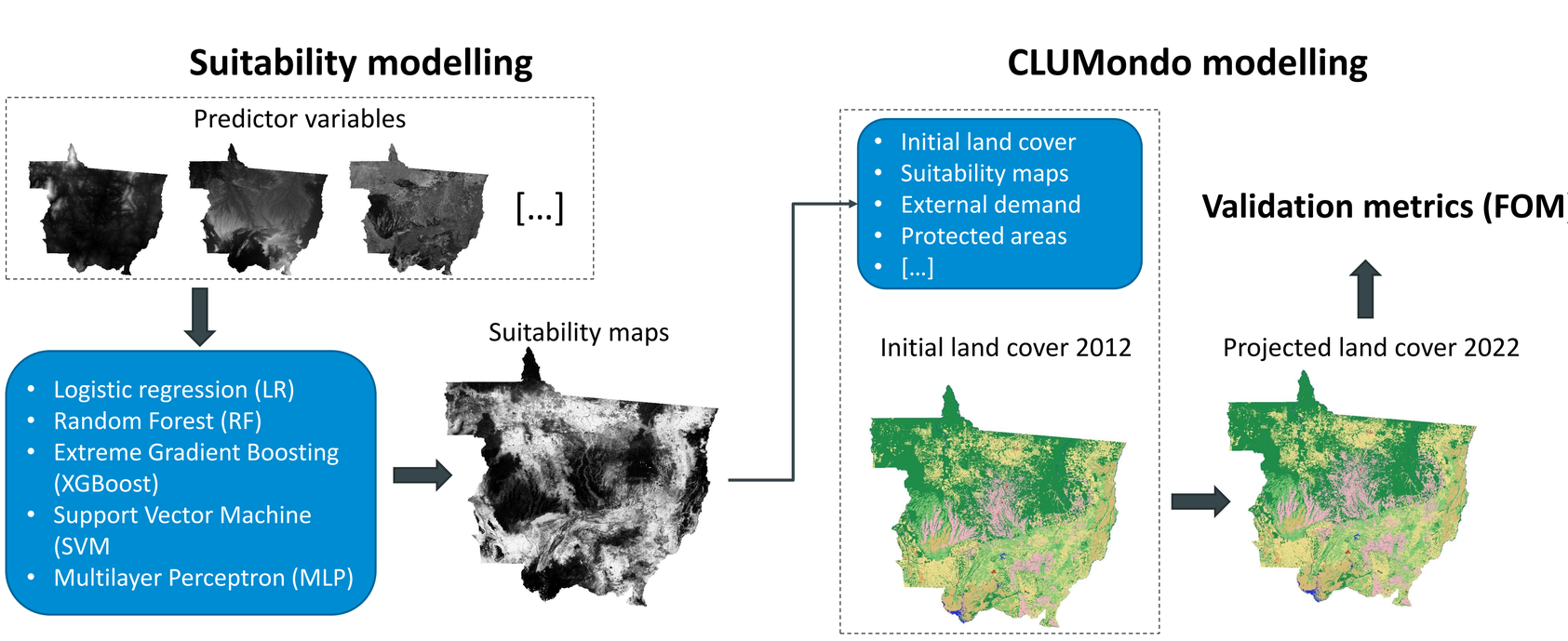
- The future is inherently uncertain, but land use models can help to envision plausible future land use scenarios
- Many land use modelling approaches rely on suitabiliy maps as input which requires geostatistical approaches to parameterise
- CLUMondo is a prominent example for a land use model, which uses logistic regression for sutability calculation in its original version
- Advanced machine learning (ML) methods have proven to ourperform logistic regression approaches in sutiability mapping

Research question:

- How much does the implementation of machine learning algorithms in CLUMondo improve modelling performance and results?

Methods:

- State of Mato Grosso (Brazil) selected as study area (903,357 km² or approx. 2.5 times the size of Germany)
- We used 22 predictor variables (environmental and socioeconomic) to derive relevant suitability maps
- We employed 5 different ML modelling methods (see for details below)
- Resulting suitabiliy maps from different methods were ingested into CLUMondo
- Projected land cover maps compared with validation metrics (Figure of Merit = FOM)



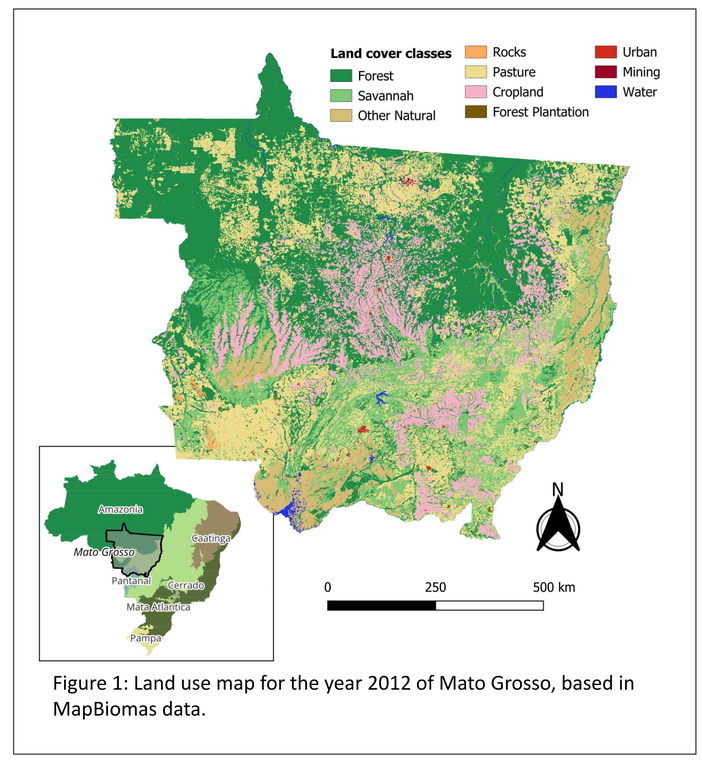
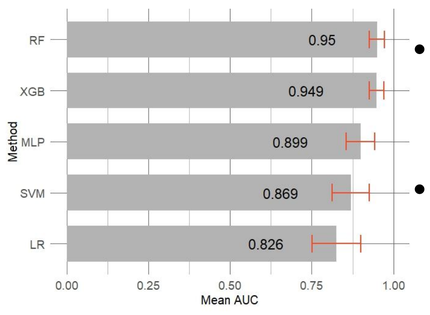


Figure 1: Land use map for the year 2012 of Mato Grosso, based in MapBiomas data.

Results:

AUC results (suitability)



- Ensemble tree models yield highest AUC
- Logistic regression lowest AUC

Figure 2: Mean AUC and standard deviation per modelling method

CLUMondo validation results

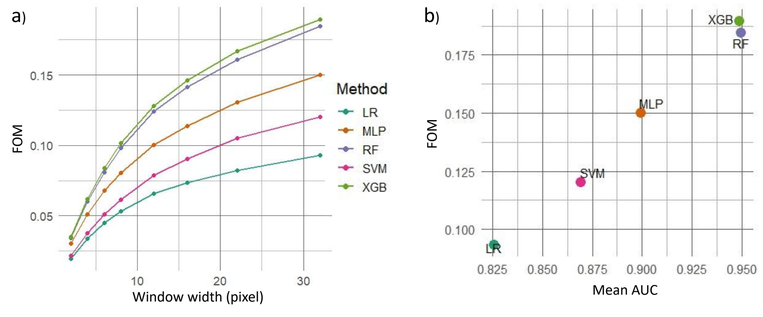


Figure 3: FOM values depending on aggregation level and suitability map input (a); Mean AUC values of suitability maos against FOM values of derived land use predictions (b)

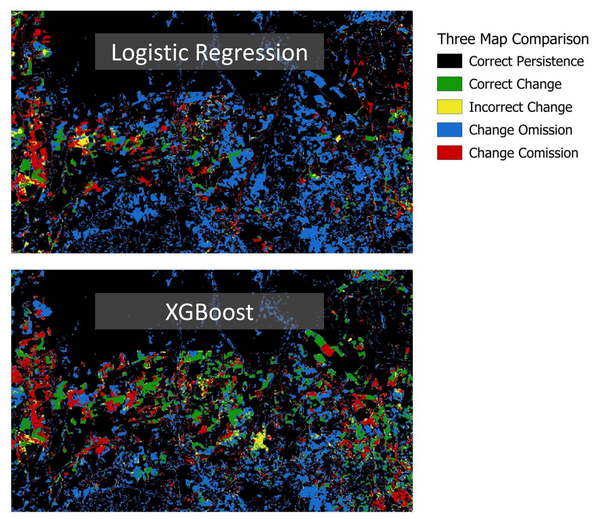


Figure 4: Example of land use allocation predictions 2022 with LR (top) and XGBoost (bottom) against observed 2022 land use maps

- High AUC values from suitability modelling translate to high FOM values in land cover projections
- Ensemble tree models (RF and XGBoost) performed best for suitability modelling and their derived suitability maps yielded highest allocation accuracy (FOM)

Discussion and implication:

- Improving quality of suitability maps contributed to overall improved future land use allocation in case of CLUMondo
- Ensemble tree models (RF, XGB) performed best in term of AUC and FOM compared to LR: 0.95 and 0.94 vs. 0.82
- Improved land use modeling performance and transparency increases credibility for future predictions and supports decision making
- Choice of model user for suitability calculation method can affect outcomes of impact assessment, e.g. changes in above-ground carbon or landscape connectivity

CLUMondoPy
GitHub repository

Funded by Federal Ministry of Agriculture, Food and Regional Identify (BMLEH)
Contact person: simon.thomsen@thuenen.de
https://www.thuenen.de/en/cross-institutional-projects/carbon-leak