Assessment of variables, species distribution models and expert knowledge for detecting Prosopis habitat in Turkana, Kenya.

Wai-Tim Ng[a], Markus Immitzer[a], Alexsandro Cândido de Oliveira Silva[b], Marcio Pupin Mello[c], and Clement Atzberger[a]

[a] Institute for Surveying, Remote Sensing and Land Information (IVFL), University of Natural Resources and Life Sciences (BOKU), Vienna, Peter Jordan Straße 82, A-1190 Vienna, Austria;
[b] National Institute for Space Research (INPE) Image Processing Division (DPI) Avenida dos Astronautas 1758 – 12201-010 – São José dos Campos, SP, Brazil
[c] The Boeing Company Boeing Research & Technology – Brazil (BR&TB) Estrada Dr Altino Bondesan 500 – 12247-016 – São José dos Campos, SP, Brazil

Introduction

Prosopis ssp., a mesquite native to dry zones in the Americas were introduced to arid and semi-arid environments for its drought tolerance and rapid growth. In Turkana the species was propagated in the 1970's for:

- stabilization of dune systems;
- provision of fuel wood; and
- restoration of degraded ecosystems.

Figure 1. Prosopis invading a riverbed.

In East Africa a number of introduced *Prosopis spp*. have hybridized and naturalized (Figure 1), becoming an aggressive invader outcompeting and replacing endemic species. The invasion is mitigated though a number of adaptations:

- the ability to produce a large number of edible and resilient seeds;
- developing an extensive root system tapping deep into the groundwater table;
- the capacity of rapid growth rates and ability to coppice after damage.
- displaying a high tolerance to climate extremes and various soil types, and having allelopathic and allelochemical effects on other plants.

The **challenges** of species distribution modelling (SDM) for invasive species are:

- organisms are not at equilibrium within their environment, and
- species absence data are often unavailable or difficult to interpret

The **aim** of the research is to:

- determine best environmental variables: synthesise expert knowledge and translate into a set of features, while determining which variables contribute to the SDM's;
- evaluate SDM's: apply and test four different SDM's for predicting habitat of invasive *Prosopis spp.*
- assess accuracy: based on AUC, Cohen's kappa, and TSS;
- conclude on potential habitat: compare mean modelling result to extent of the *Prosopis* cover for Turkana, Kenya (Figure 2, Ng et al. 2016).

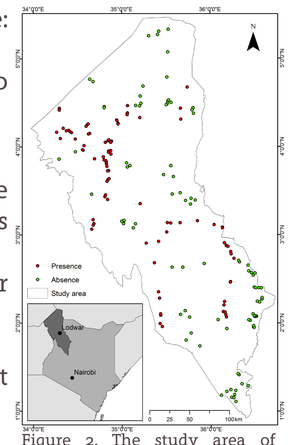


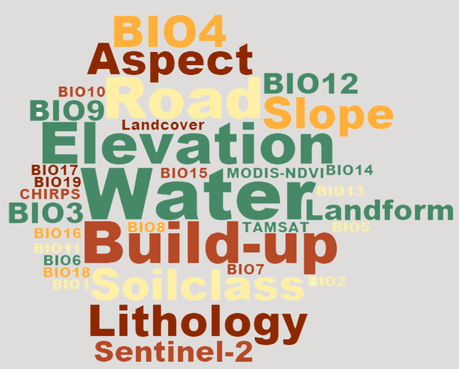
Figure 2. The study area of Turkana, kenya and the absence and presence points.

Material and Methods

A. Environmental variables and expert knowledge

We collected a set of 33 environmental variables. These were interpreted based on expert knowledge and a literature review, then gradually reduced to the eight features by:

- variable importance, as determined by the variables contribution and jack-knife test;
- variable multicollinearity, as determined by a pair-wise Pearson and Spearman correlation tests; and
- variable bias, as determined by assessing the outputs and identifying overly dominant variables.



B. Species Distribution Models

We selected and assessed four models, ranking from fundamental and widely used SDM's, Logistic regression (LR) and Maximum entropy (ME), to more advanced and innovative SDM's, Random Forest (RF) and Bayesian Networks (BN) (Silva et al. 2014).

C. Model Validation

The SDM's were evaluated based on three independent tests:

- area under the receiver operating characteristic (ROC) curve (AUC);
- Cohen's kappa; and
- true skill statistics (TSS).

Results and Discussion

A. Variable importance and selection

We determined that following eight features were best suited for modelling potential *Prosopis* habitat: distance to water, built-up and roads, lithology, dominant soil type, landform, elevation, and temperature seasonality (the difference between the annual maximum and minimum temperatures, Figure 3)

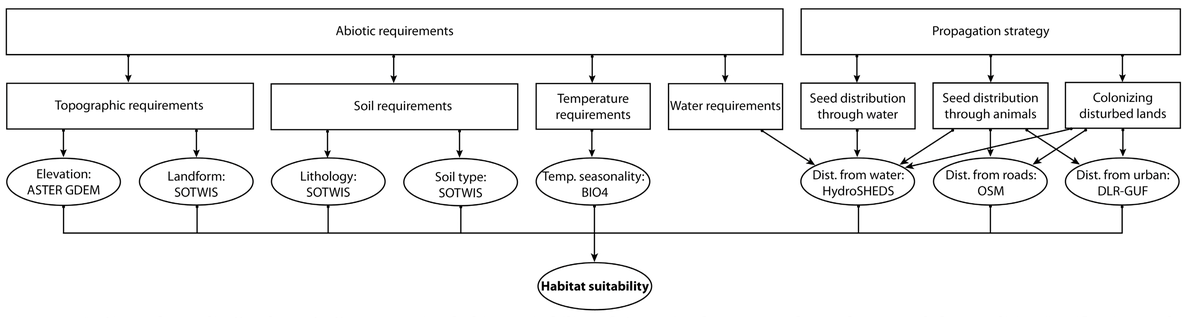


Figure 3. A Directed Acyclic Graph or DAG representing habitat suitability of Prosopis. The rectangular nodes proved the condition/justification and underlying process for using a variable.

B. Model outputs

The models used identical sets of environmental variables and produced outputs which we compared after discretization into four classes (Figure 5, top). The default value is arbitrarily set at 0.5. Unsuitable habitat is characterized by a denominator close to 0, while suitable areas have a value close to 1.

We created a habitat suitablility map by averaging the four model outputs of each tested SDM (Figure 5 , bottom).

Figure 6 overlays the *Prosopis* cover, derived from a Random Forest classification using Sentinel-2 data, with the mean modelling result displaying good overlap between the modelling results and land cover classification.

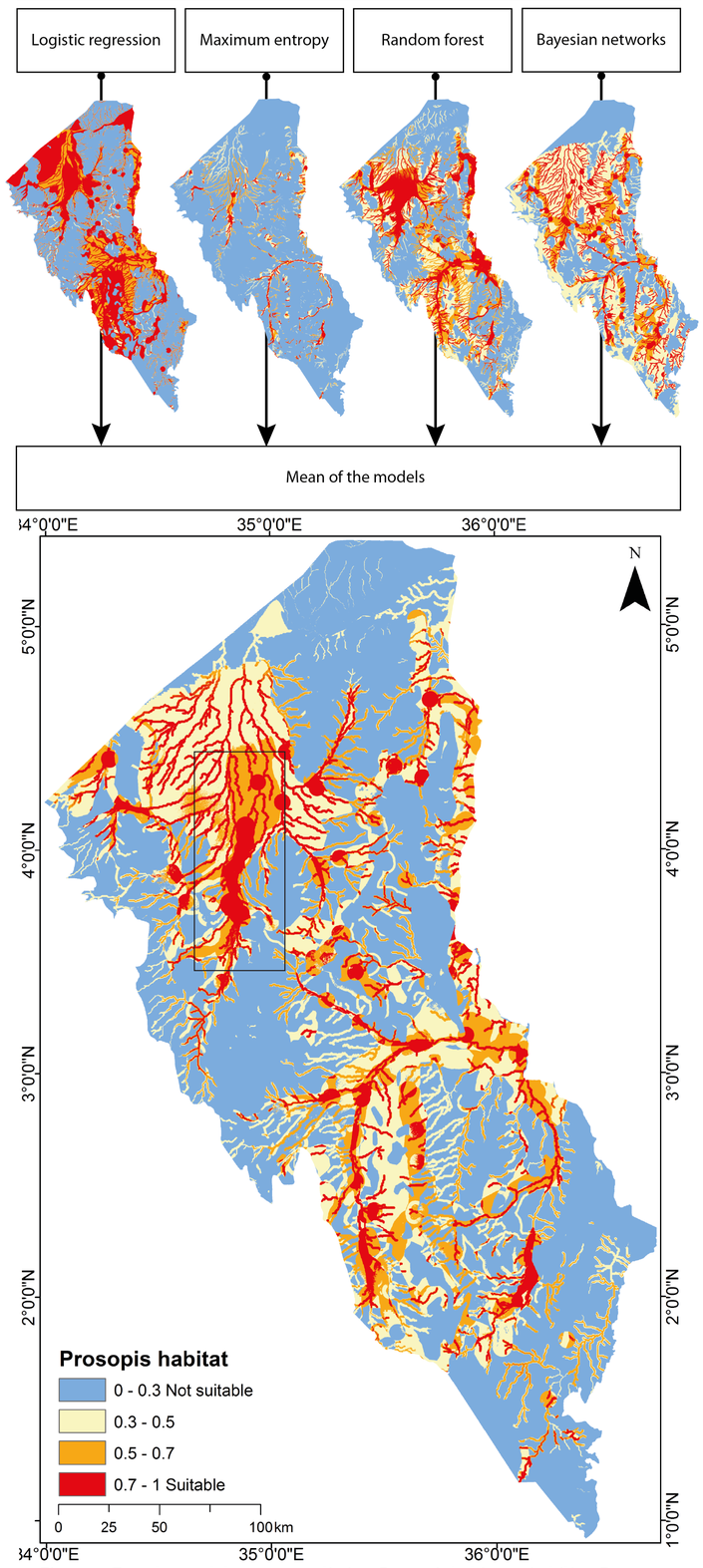


Figure 5. Top, the model output predicting *Prosopis* habitat suitability. Bottom, the mean model output and the extent of Figure 6 (black box).

C. Accuracy assessment

We compared and assessed the model outputs at Table 1. ROC/AUC, Cohen's kappa and TSS were generated from confusion matrices, which were compiled from the reference dataset, consisting of presence and absence data, and the extracted and dichotomized values (0 or 1) of the model outputs.

Figure 4 displays the values of the predicted results for each model and the reference data (absence and presence).

Table 1. Accuracy assessment of the modelling results

	ROC/AUC	Cohen's kappa	TSS
LR	0.914	0.8252	0.8255
ME	0.883	0.6935	0.6923
RF	0.940	0.8798	0.8799
BN	0.924	0.8470	0.8468

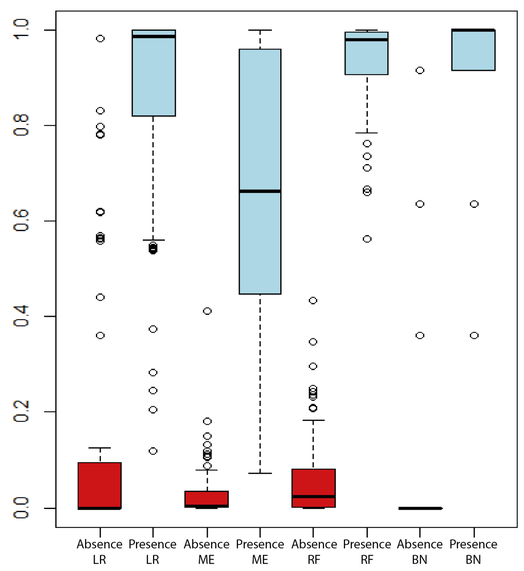


Figure 4. Boxplots of the reference data and the raw prediction value received from the model.

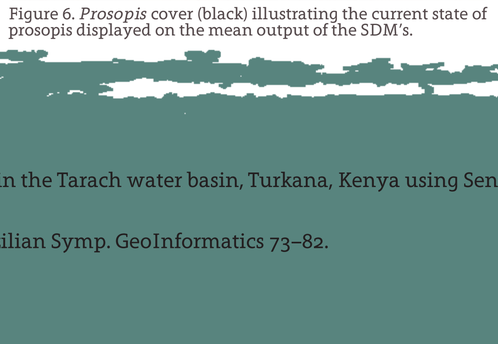


Figure 6. *Prosopis* cover (black) illustrating the current state of prosopis displayed on the mean output of the SDM's.

Conclusions

We can conclude that:

- Expert knowledge provided the groundwork for our analysis and had a positive effect on the results;
- driving factors proved to be: distance from water, urban centres and roads, soil type, lithology, landform, elevation and temperature seasonality;
- Random Forest and Bayesian network models provided highest accuracies and most plausible results;
- The invasion pattern is in line with literature and the models indicate that high risk areas correlate with ecologic and economical valuable and vulnerable areas. Despite being moderate in size they have a large impact on livelihoods and biodiversity.

References

Ng, WT; Immitzer, M; Floriansitz, M ; Vuolo, F; Luminari, L; Adede, C; Wahome, R; Atzberger, C (2016) Mapping Prosopis spp. within the Tarach water basin, Turkana, Kenya using Sentinel-2 imagery. SPIE 9998, Remote Sensing for Agriculture, Ecosystems, and Hydrology XVIII, 99980L, doi: 10.1117/12.2241279

Silva, A.C.D.O., Mello, M.P., Fonseca, L.M.G., (2014) Enhancements to the Bayesian Network for Raster Data (BayNeRD). Proc. Brazilian Symp. GeoInformatics 73–82.

Contact: tim.ng@boku.ac.at

University of Natural Resources and Life Sciences, Vienna

EMAPI 14
International Conference on Ecology and Management of Alien Plant Invasions
PORTUGAL | 2017