Essential environmental variables to include in a stratified sampling design for a national-level invasive alien tree survey

Johann DF Kotze (1-2), Hein B Beukes (1), Thomas Seifert (2-3)

Introduction

Alien plant invasions are known to have severe disruptive impacts on biodiversity, ecosystems, plant and animal populations, ecosystem services, agriculture, forestry, the economy and human welfare (Jeschke et al. 2014, Vilà & Hulme 2017). One of the most important attributes of biological invasions in terms of impact is invasive species abundance (Kumschick et al. 2015). In other words, the more there is of an invasive alien plant (IAP) species, whether number of individual plants or biomass, the greater the impact. Thus in particular invasive tree species with their large biomass and their ability to change their local environment substantially have an impact on ecosystems and ecosystem services (Le Maitre et al. 2016). Mitigation strategies to deal with alien plant invasions have been implemented across the world with noted successes in the control of invasive species (Simberloff et al. 2011). In South Africa, the long-term Working for Water Programme was initiated in 1996 as an IAP control programme sponsored by government (Van Wilgen et al. 2012). South Africa, with its rich biodiversity (Driver et al. 2012), has been invaded by many different IAP species, especially tree species ( Nel et al. 2004), and the ecological and economic impacts of these invasions have been well documented (De Lange & Van Wilgen 2010, Le Maitre et al. 2016). Further to this, South Africa hosts three of the 35 current biodiversity hotspots in the world (Mittermeier et al. 2011). This makes the threat of IAP species to this region an international concern (Mitter-
meier et al. 2011). Species abundance data is essential in the effective management of such control programmes and serves as an important indicator in the measurement of their success (Wilson et al. 2018). To priori-
tise intervention or mitigation strategies at a national level, it is important to achieve IAP distribution and abundance data at this scale. Despite the success of many of these initiatives, they still lack sound protocols for objectively determining IAP distribution at a national level, with the obvious result of not being able to measure and monitor abundance or spatial extent and abundance changes over time (Dehnen-Schmutz et al. 2018).

The spatial extent of the study area (South Africa covers approximately 122 mil-

lion hectares), the environmental and eco-

logical heterogeneity (Driver et al. 2012), and

limited resources to conduct surveys (Ricciardi et al. 2017), make a complete IAP inventory not feasible to carry out (Web-

ster & Lark 2013). The best alternative for providing unbiased and reliable quantita-

tive information is a partial estimation based on sampling (Gitzen et al. 2012). An example of this is the statistical or sample based surveys which have been applied for many years in most large scale forestry sur-

veys in many countries (Stahl et al. 2016), and are based on strict design-based princi-

ples (Naesset et al. 2017). The success of such monitoring programmes is deter-

mined by the underlying sampling design (Gitzen et al. 2012). Ideally, the sampling strategy should effectively represent the variability of the entire target population with as few as possible sample points. The simple random sampling design is known to be inefficient in providing an even repre-

sentative coverage of a study area, due to the tendency of sample point locations to cluster at low sampling intensities, result-

ing in large undetected areas (Webster & Lark 2013). The result is that resource de-
mands such as costs, manpower and time

required for random sampling designs are high (Kalkhan 2011) if the aim is to ensure that the inherent variation in the target population is represented (Webster & Lark 2013). An alternative is to use a pre-strati-

fied sampling design which improves the accuracy of the estimates and allows for a better efficiency (Webster & Lark 2013).

The objective of stratification for vegeta-
tion surveys is to incorporate those habitat types that show the most meaningful asso-

ciation with the vegetation attribute of in-

terest, and so to ensure that all the possi-

ble habitat specific variation that contrib-

utes to the target species range and abun-

dance is included in the survey (Gitzen et al. 2012). The latter also provides well-de-

fined strata, which allows for effective comparisons across strata for valid infer-

ence between field observations (Webster & Lark 2013). The challenge in this context is the selection of environmental variables which adequately define spatial units rep-

resenting homogenous conditions or stra-
ta for species abundance. These strata should clearly reflect the relationship be-

 tween IAP species and their physical envi-

 ronment and thereby summarize this un-

derlying non-random relationship (Vollis 2016). The main aim of stratification is thus to minimise the variance within the strata while maximising the variance between them. All of which leads to the main ques-
tion: which predictive environmental vari-

ables and how many of them should be in-

 cluded for defining the strata while main-

 taining a full rank design. Such a design provides the most effective inference be-

 tween species’ observations obtained from

actual field surveys.

Appropriate methods to model the corre-

lation between species’ occurrence and en-

vironmental variables such as climate, soil

to and terrain are predictive vegetation or

species distribution models (SDM – Hageer et al. 2017). SDMs not only provide insights into the species-environment relationships, but they are also used to predict spatial dis-

 tributions of target species by means of

maps of the correlated environmental pre-
dictor variables (Elith & Franklin 2017). Mul-
tiple ways have been proposed to model

species distribution and prominent exam-

ples include regression trees, boosted re-

gression trees and random forests machine

learning algorithms that are used to com-

bine rules for species occurrence in an opti-

 mum way (Franklin 2010). Examples for

rule-based systems are GARP (Stockwell 1999) that applies a genetic algorithm or MaxEnt (Anderson et al. 2003) that works on a maximum entropy optimisation. Other authors applied a traditional parametric al-

 gorithm such as regression analysis (Fahnmeir et al. 2013). Most methods are

known to provide equally good results (Aitor & Garcia-Viñas 2011, Sahragard &

Ajorloa 2018). Species distribution model-

ling has been widely applied in the field of

invasion biology for a range of objectives (see Robinson et al. 2017 for a review). For

instance, Rouget et al. (2015) used broad scale predictor variables that included cli-

 mate, natural biomes and anthropogenic

 factors in relationship to the distribution of IAP species’ assemblages in an effort to

 map wide-ranging alien plant biomes. Ap-

 plications also include the use of models to support the development of appropriate

sampling designs and includes the defini-

 tion of appropriate strata (Särndal 2010).

In this study we assessed the extent to which the modelled associations between IAP species and environmental variables were meaningful based on repeated corre-

 lation patterns across extensive areas with high levels of environmental variation. We hypothesised that, although localized associa-

tions between IAP species and different environmental variables might vary, there would be constant regional correlation pat-

 terns with a limited number of specific vari-

abies.

The objective of the stratification process was to obtain all possible combinations or inter-

 actions between the different environ-

 mental variables. This full rank design is ad-

 vantageous from a statistical point of view

 and easily obtained in a controlled environ-

 ment or a planned experiment. The chal-

 lenge is to obtain such a design within the

 natural environment at a national scale. Thus the aim of this study was to combine the unique and varying natural geographi-

cal distribution patterns of the underlying
deterministic environmental variables to
effectively summarise IAP species’ abun-
dance within these strata, whilst maintain-

 ing a complete full rank design.

This paper presents an approach to firstly filter and select environmental variables most suitable for such a national-level de-

 sign-based stratification in South Africa and, secondly, to explore how many catego-

 ries could realistically be included in such a stratification. The results represent the first phase in establishing a regional level IAP monitoring programme for South Africa on a scientific and statistically rigor-

ous basis.

Materials and methods

Study area

The study area was the whole of South Af-

 rica and the focus was on undisturbed areas or rather natural and semi-natural ar-

 eas or habitats as defined by Nel et al. (2004), namely: “natural and semi-natural ecosystems, that is, those that are still rea-

sonably intact, having most of their biodi-

 versity structure and functioning, and with

 primary driving forces operating within natural/evolutionary limits”. These habitats are most threatened by IAP species by hav-

 ing the greatest impact on native biodi-

 versity and ecosystem services (Nel et al. 2004).

IAP distribution records

The most comprehensive set of records of the spatial distribution of IAP species for the study area is the Southern African Plant Invaders Atlas (SAPIA) database that con-

 tains records for more than 500 different IAP species (Henderson & Wilson 2017). SAPIA observed IAP species with no under-

 lying statistical basis along road transects of 5-10 km long and within the adjacent road area from a moving vehicle (Hender-

son & Wilson 2017). IAP species were most-

 ly recorded per quarter degree square (QDS), a 15’ latitude × 15’ longitude square, therefore the exact location of species was related to a total area of approximately 25-27 km or 65,000 ha. As many as 120 dif-

ferent IAP species were recorded per QDS and often with repetitive observations per species. An abundance value is provided for each record based on the approximate number of actual plants observed per unique IAP species within a 10 km transect. A number of habitat classes are also pro-

 vided per species record to allow for a species to be classified based on habitat preference. Many of these species have a
limited distribution and abundance, overlap in distribution and are biased towards certain habitat classes, so the SAPIA database was filtered for the study using a stepwise rule-based approach (for further details on the species filter process, see Appendix 1 and Fig. S1 in Supplementary material). Species were firstly selected on the basis of having the maximum distribution range across South Africa at a high abundance. This captured the full environmental gradient contributing to a particular IAP species’ observed distribution. Subsequently, species with minimum overlap in spatial extent with other species were identified to create mutually exclusive observations for each of the IAP species across geographic space. The combination of maximum spatial distribution with minimum overlap led to the selection of three tree species, namely *Acacia cyclops*, *Acacia mearnsii* and *Prosopis glandulosa* (Fig. 1).

**Environmental variables**

A set of physiologically relevant environmental variables that have been shown to correlate with species abundance were included, namely climatic, topographic (terrain) and soil related variables (Williams et al. 2012, Hageer et al. 2017, Fois et al. 2018). The climatic variables were obtained from the WorldClim2 dataset (Fick & Hijmans 2017). Soil variables were extracted from the South African Land Type Survey database, which is based on detailed field surveys published at a 1:250,000 scale (Land Type Survey Staff 2006). Terrain variables such as aspect were derived from the Shuttle Radar Topographic Mission (SRTM) digital elevation data at the 90 m resolution (Farr et al. 2007).

Environmental variables were resampled to a 400 × 400 m spatial resolution where required (Tab. 1). Multicollinearity (Franklin 2010) amongst predictor variables was assessed by means of the pair-wise correlation coefficient between variables (Williams et al. 2012). Pairwise correlation exceeding a threshold collinearity of more than 0.75 (Dormann et al. 2013) was used to exclude variables.

**Spatial combination of species presence with environmental variables**

Each of the three species’ layers was spatially intersected with the overlapping environmental variable matrices to create three unique species/environmental datasets by means of ArcGIS® Desktop software (ESRI 2017). The application of the South African tertiary catchment delineation as an aggregation unit supplied replications across geographic space for each of these three layers. Catchment delineation was applied as an aggregation unit for it is defined by topography. Catchments therefore captures the full range of terrain morphological units which ensures that soil and climatic gradients are included. Adjacent catchments have closely matching gradients of these variables which makes them reasonable replicates for determining strata. Further to this, catchment delin-

---

**Fig. 1** - A description and distribution of the three identified species, namely *Acacia cyclops*, *Acacia mearnsii* and *Prosopis glandulosa*.

- **Acacia cyclops**: Introduced from Australia, it forms dense evergreen shrubs to an eight metre tall tree with an average height of three metres. Plants carry lots of dead wood and old seed pods. It grows mostly in the winter rainfall areas or areas that receives rainfall throughout the year and typically below 300 metres above sea level. This species occurs on acidic and calcareous sands and often dominates low lying coastal flats.

- **Acacia mearnsii**: This evergreen tree originating from Australia reaches heights of between 10 to 30 metres. Trees carry a fair amount of old wood and pods as well as large amounts of conspicuous pale yellow flowers during late winter and early spring. It occurs from sea level to high altitude areas across a broad rainfall region. Fairly drought and frost resistant, but prefers more temperate climates and lower lying valley bottoms. It occurs over a broad spectrum of soils.

- **Prosopis glandulosa**: This species originates from the South-western United States and Northern Mexico. It typically occurs in the more arid regions of South Africa. Favours habitats with deep soils and where ground-water is available which includes riparian zones, seasonal watercourses, pans and depressions. It usually grows as multi-stemmed shrubs to small trees averaging from 2 to 4 metres. Can tolerate extended drought periods.
Tab. 1 - Environmental variables used in the analysis.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Resolution (m)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Climate</td>
<td>Annual Mean Temperature</td>
<td>1000 × 1000</td>
<td>Fick &amp; Hijmans 2017</td>
</tr>
<tr>
<td></td>
<td>Mean Diurnal Range (Mean of monthly [Max Temp - Min Temp])</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Isotemperality (Mean Diurnal Range / Temperature Annual Range) (&lt;100)</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temperature Seasonality (standard deviation ×100)</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max Temperature of Warmest Month</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min Temperature of Coldest Month</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temperature Annual Range (Max Temperature of Warmest Month - Min Temperature of Coldest Month)</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean Temperature of Wettest Quarter</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean Temperature of Driest Quarter</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean Temperature of Coldest Quarter</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Annual Precipitation</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precipitation of Wettest Month</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precipitation of Driest Month</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precipitation Seasonality (Coefficient of Variation)</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precipitation of Wettest Quarter</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precipitation of Driest Quarter</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precipitation of Coldest Quarter</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precipitation of Wettest Quarter</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precipitation of Driest Quarter</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precipitation of Coldest Quarter</td>
<td>1000 × 1000</td>
<td></td>
</tr>
<tr>
<td>Soil</td>
<td>Soil depth (mm)</td>
<td>400 × 400</td>
<td>Land Type Survey Staff 2006</td>
</tr>
<tr>
<td></td>
<td>Percentage clay in the A-horizon</td>
<td>400 × 400</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Percentage clay in the B-horizon</td>
<td>400 × 400</td>
<td></td>
</tr>
<tr>
<td>Terrain</td>
<td>Terrain morphological units (valley bottom, footslope, midslope, scarp and crest)</td>
<td>90 × 90</td>
<td>Land Type Survey Staff 2006</td>
</tr>
<tr>
<td></td>
<td>Elevation (m a.s.l.)</td>
<td>90 × 90</td>
<td>Farr et al. 2007</td>
</tr>
<tr>
<td></td>
<td>Aspect</td>
<td>90 × 90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slope (%)</td>
<td>90 × 90</td>
<td></td>
</tr>
</tbody>
</table>

Environmental modelling

Species abundance served as the response variable, whilst the environmental variables were applied as predictor variables. Relationships between the response variables and each predictor variable were investigated by means of visual inspection of the resulting graphs to determine the type of correlation models to use from the wide range of techniques available for modelling species-environment associations. Data distribution guides the selection of the type of modelling approach to apply (Dormann 2011). The relationships were linear, resulting in opting for a more traditional modelling approach, namely linear regression models (Dormann 2011). These were developed based on the generalized linear model (GLM) framework (Fahrmeir et al. 2013). GLM's are extensively applied in species distribution modelling due to their strong statistical foundation and ability to realistically model species-environment associations (Elith & Franklin 2011).

Data outliers were identified and subsequently removed for each environmental variable per tertiary catchment based on set cut-off limits applied to the left and right of the normal distributions per variable. Cut-off limits were based on the variable's coefficient of variation (CV). For instance, for a CV<10 the applied cut-off z-value is 1.96, whilst for a CV>20 the applied cut-off z-value was decreased to 1.65.

The three IAP species were investigated and analysed independently per tertiary catchment. Statistical analysis was conducted by means of Matlab® software (MathWorks 2017). The environmental predictor variables were added one at a time to the model, and all possible combinations of the number and type of variables were explored for their effect on model performance. This resulted in a multitude of models per tertiary catchment for each species. Akaike’s Information Criterion (AIC) was used to evaluate models per catchment and select the most appropriate model with its associated predictor environmental variables (Symonds & Moussalli 2011).

Stratification simulation

Environmental variables were reclassified into three classes each based on a gradient ranging from low to medium and finally high for two aggregation levels or spatial scales, namely the complete study area (South Africa) and the tertiary catchment delineation. Interaction classes between variables were created by intersecting the different variables in geographic space for these two aggregation levels by progressively increasing the number of environmental variables in subsequent intersections (see Appendix 2 in Supplementary material for an explanation on the stratification procedures). The number of interaction classes created at the two aggregation levels at each intersection level were compared with the number of classes required for an ideal theoretical full rank design. This comparison provided an indication of an appropriate number of variables to be included in such a stratification exercise to achieve a design as close as possible to, if not a complete factorial design.

The three IAP species were then combined with the created strata at the maximum identified intersection level before actual stratification started to deviate from
the ideal full rank design. The effectiveness of the identified environmental variables and the subsequent stratification to reduce the variation of IAP distribution and abundance within strata was compared to the overall variation without any form of stratification at a tertiary catchment level. Should the stratification be meaningful, IAP abundance variation at a stratum level would be significantly less than at an un-stratified level on a repetitive basis across tertiary catchments. An analysis of variance (ANOVA) was then applied to the data to simulate a data analysis using the data as if it came from an actual IAP survey to see if there was a significant association between IAP distributions and the respective strata. Should there be no association between IAP distribution and respective strata, therefore IAP abundance varied at random across strata, the use of strata as categories to describe IAP distributions as a response variable within them would be meaningless.

Results

Environmental modelling

Fourteen variables from the original 26 remained after testing for multicollinearity among predictors. A threshold was applied to select environmental variables most frequently associated with the three IAP species, namely those variables repetitively observed more than 75% per species across tertiary catchments. In the case of A. cyclops these variables included soil depth, percentage clay in the A-horizon, percentage clay in the B-horizon, slope and the terrain morphological units. Variables most frequently associated with A. mearnsii were the terrain morphological units, percentage clay in the A-horizon, percentage clay in the B-horizon, soil depth, long-term mean annual precipitation and isothermality. P. glandulosa was mostly associated with clay in the B-horizon, soil depth and long-term mean annual precipitation (Tab. 2). The total percentage association between environmental variables and the three species combined was then determined to provide an overall indication of association per variable across all species. (Fig. 2). Further filtering of variables was based on a combination of ecological reasoning (Dormann et al. 2013) and the frequency of occurrence of variables for all three IAP species.

Stratification simulation

The stratification of the complete study area up to the spatial intersection of five environmental variables generated 243 classes, which was similar to the total number of possible classes at that level for a full rank design within a controlled experiment (Fig. 3). Stratification at the smaller aggregation level, namely the tertiary catchment delineation, started to deviate from a full rank design with the intersection of between three and four variables, therefore between 27 and 81 classes. When this was done with five or more variables with three levels each, the stratification at a tertiary catchment level started to deviate substantially from the total amount of all possible class combinations (Fig. 3). The

<table>
<thead>
<tr>
<th>Tab. 2 - Environmental variables associated the most frequently with the different species (&gt;75%).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental Variables</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>Annual precipitation</td>
</tr>
<tr>
<td>Percentage clay in the A-horizon</td>
</tr>
<tr>
<td>Percentage clay in the B-horizon</td>
</tr>
<tr>
<td>Soil depth</td>
</tr>
<tr>
<td>Isothermality</td>
</tr>
<tr>
<td>Slope</td>
</tr>
<tr>
<td>Terrain morphological units</td>
</tr>
</tbody>
</table>

![Fig. 2](image2.png) - The total percentage association between the specific predictor environmental variables and the three tree species combined for all tertiary catchments.

![Fig. 3](image3.png) - The number of unique strata created by means of intersecting environmental variables. The graph only includes up to the intersection of seven variables with three even area classes each for thereafter the difference in number of obtained intersection classes only increases.
Further testing of the feasibility of the stratification was based on a stratification done at a tertiary catchment level by including four variables with three levels each and thereby 81 possible unique interaction combinations or strata per catchment. The variance in IAP abundance per stratified tertiary catchment was significantly lower than the variance in related tertiary catchments without any stratification across all tertiary catchments (Fig. 4), indicating that stratification had a substantial effect. The results of the analysis of variance applied to the same dataset showed significant differences in mean IAP abundance variation as summarized by the stratification for each of the three species (Tab. 3, Tab. 4).

**Discussion**

The results of this study are a method for stratified sampling as a base for a large scale inventory of invasive tree species or other invasive alien plants at a national level. The proposed coherent and objective method provides a means to use edaphic, climatic and geomorphologic variables to choose adequate strata in order to gain the necessary sampling efficiency for larger areas. It closes an obvious gap in IAP monitoring, where the current methods lack a statistical rigorous design-based approach and have mainly relied on either opportunistic recording of IAPs along accessible pathways such as roads (Hansson & Wilson 2017) or have been used to get presence/absence information based on expert knowledge, literature and herbarium records (Vinogradova et al. 2018). Although species distribution modelling is a standard tool to predict potential IAP distribution (Robinson et al. 2017), the objective of this study was not to map potential

---

**Tab. 3** - ANOVA summary including main effects and the intersection of the predictor environmental variables up to the 1st order. IAP species abundance served as the response variable (level of significance applied was p<0.05). (df): degrees of freedom.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Sum of squares</th>
<th>df</th>
<th>Mean square</th>
<th>F-value</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>10211609823</td>
<td>1</td>
<td>1021609823</td>
<td>1183.68</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rainfall</td>
<td>776469154</td>
<td>2</td>
<td>388234577</td>
<td>451.80</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Soil Depth</td>
<td>1680298283</td>
<td>2</td>
<td>840149141</td>
<td>977.72</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Clay B-hor</td>
<td>155313015</td>
<td>2</td>
<td>77656085</td>
<td>90.37</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Terrain morphology</td>
<td>1602866749</td>
<td>2</td>
<td>801433375</td>
<td>932.66</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rainfall × Soil depth</td>
<td>1538399235</td>
<td>4</td>
<td>384599809</td>
<td>447.57</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rainfall × Clay B-hor</td>
<td>232806813</td>
<td>4</td>
<td>58201703</td>
<td>67.73</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rainfall × Terrain morphology</td>
<td>313297417</td>
<td>4</td>
<td>783243579</td>
<td>911.49</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Soil depth × Clay B-hor</td>
<td>444560606</td>
<td>4</td>
<td>111140151</td>
<td>129.34</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Soil depth × Terrain morphology</td>
<td>259746046</td>
<td>4</td>
<td>64936512</td>
<td>75.57</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Clay B-hor × Terrain morphology</td>
<td>1227720920</td>
<td>4</td>
<td>306930230</td>
<td>357.19</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Error</td>
<td>9118077400</td>
<td>10611</td>
<td>859297</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Tab. 4** - ANOVA table with all possible levels of intersection up to the 3rd order. IAP species abundance served as response variable and the environmental variables as predictor variables (level of significance applied was p<0.05). (df): degrees of freedom.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Sum of squares</th>
<th>df</th>
<th>Mean square</th>
<th>F-value</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept × Soil depth</td>
<td>1042058127</td>
<td>1</td>
<td>1042058127</td>
<td>12306.99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rainfall × Soil depth</td>
<td>570818379</td>
<td>4</td>
<td>142704595</td>
<td>168.54</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rainfall × Clay B-hor</td>
<td>526171140</td>
<td>4</td>
<td>131542785</td>
<td>155.36</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Soil depth × Clay B-hor</td>
<td>458571226</td>
<td>4</td>
<td>114642806</td>
<td>135.40</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rainfall × Terrain morphology</td>
<td>299307134</td>
<td>4</td>
<td>748254336</td>
<td>883.71</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Soil depth × Terrain morphology</td>
<td>808312529</td>
<td>4</td>
<td>202078132</td>
<td>238.66</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Clay B-hor × Terrain morphology</td>
<td>959066426</td>
<td>4</td>
<td>239766606</td>
<td>283.17</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rainfall × Soil depth × Clay B-hor</td>
<td>109635396</td>
<td>8</td>
<td>137066925</td>
<td>161.88</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rainfall × Soil depth × Terrain morphology</td>
<td>949042946</td>
<td>8</td>
<td>118630368</td>
<td>140.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rainfall × Clay B-hor × Terrain morphology</td>
<td>884723287</td>
<td>8</td>
<td>110590411</td>
<td>130.61</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Soil depth × Clay B-hor × Terrain morphology</td>
<td>504441130</td>
<td>8</td>
<td>63055141</td>
<td>74.47</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rainfall × Soil depth × Clay B-hor × Terrain morphology</td>
<td>244633516</td>
<td>16</td>
<td>152895845</td>
<td>180.57</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Error</td>
<td>89812467726</td>
<td>106071</td>
<td>846720</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
IAP distribution, but rather to use modelling to support the development of a stratification that could be used in a sampling design (Soti et al. 2015) in order to quantify IAP abundance based on a representative grid of empirical sampling points. Similar approaches have been used to guide surveys for example where field surveys are limited due to a lack of resources (Fois et al. 2018). In these cases post-model field surveys were targeted on where a high probability of occurrence was predicted but without pre-model field data (Peterman et al. 2010). In other studies, this approach has been used to improve the assessment and verification of the distribution of scarce species and to optimise resources by focusing surveys on localities where a high probability of occurrence of such rare species was predicted (Peterman et al. 2013). This study, based on three invasive tree species of major ecological relevance, serves as the first step in the establishment of a scientifically-based regional level IAP monitoring programme for South Africa. Such a monitoring programme requires that actual IAP distribution and abundance data is sampled in the field and the resulting data should be used to iteratively refine and optimize future national level surveys (Volis 2016, Fois et al. 2018).

The results of this study revealed distinct species-specific differences in the occurrence patterns of the three invasive tree species under consideration, which may point to their ecological differences and optimum habitats. These three IAP tree species were introduced to South Africa with specific objectives and thereby established on a non-random basis. Acacia mearnsii was planted on a wide scale by the commercial forestry sector for its high tannin content in the bark. Acacia cyclops was used to stabilize drift sands along the coast and Prosopis glandulosa was planted extensively in the arid regions for animal fodder. Although all these species have had a residence time well in access of a 100 years in South Africa, it is possible that they have not reached their full geographic extent and their distribution is therefore not yet in equilibrium, which could cause problems for correlative models as pointed out by Robinson et al. 2017. This is an unknown and was mitigated for by selecting those species with the largest possible geographical extent.
population by means of grouping a continuous varying landscape into discrete classes or strata of similar variability, the sample variance was significantly reduced and sampling efficiency was increased to a level where large scale inventories are viable. The objective of this study was to determine which environmental variables most effectively summarize invasive tree abundance variability as well as to determine the number of strata to be included in such a stratification. These variables are to be applied in a future national level stratification by demarcating habitat types contributing the most to IAP occurrence in South Africa. This will ensure that all different habitat types are sufficiently included in a national level survey, as well as an optimized sample point allocation. It was shown that ideally not more than 81 unique strata should be created to obtain a stratification that does not deviate significantly from a statistically desirable full rank design. The number of variables included is obviously related to their levels and in this case it was shown that four variables at three different levels each can be used. Selected variables were identified based on a combination of correlation with species, replication across species as well as geographic space, and finally explained by means of biological reasoning. These variables included average rainfall, soil depth, clay content in the B-horizon and a form of landscape position such as terrain morphological units.

Acknowledgements

The National Invasive Alien Plant Survey (NIAPS) project is funded by the Working for Water Programme that resides with the Department of Environmental Affairs’ Natural Resource Management Programme. This work originated from the NIAPS project and the authors acknowledges the financial contribution made by the Working for Water Programme. The last author acknowledges the contribution made by the "Care4C" project, grant no. 778322, in the EU Horizon 2020 Marie Skłodowska-Curie program.

References

Peterman WE, Crawford JA, Kuhn AR (2013). Using species distribution and occupancy moditification...
Sampling design for a national-level invasive alien tree survey


Appendix 1 - Description on the approach used to filter IAP species within the SAPIA database.

Appendix 2 - The stratification procedures followed of environmental variables.

Fig. S1 - Approach followed to filter the SAPIA database.

Link: Kotze.2767@suppl001.pdf