Ensemble modeling for the potential distribution of invasive weed *Verbesina encelioides* in South Africa from 2020 to 2090

Moleseng C. Moshobane1,2,* and Luiz Fernando Esser3

1South African National Biodiversity Institute, Pretoria National Botanical garden, 2 Cussonia Avenue, Brummeria, Silverton, 0184, South Africa
2Centre for Functional Biodiversity, School of Life Sciences, University of KwaZulu-Natal, P/Bag X01, Scottsville, Pietermaritzburg, 3209, South Africa
3Laboratorio de Fitoecologia e Fitogeografia, Programa de Pós-Graduação em Bota’nica, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

*Corresponding author
E-mail: moshobanemc@gmail.com

**Abstract**

*Verbesina encelioides* (Cav.) Benth. & Hook. Fil ex Gray (Asteraceae) is a perennial herb species belonging to the family Asteraceae. The plant is an aggressive obligate outbreeder weed that has invaded vast expanses of pastures, orchards, and forest areas in tropical and subtropical regions. The purpose of this study was to determine the suitability of current and potential future habitats for *V. encelioides*, an invasive weed in South Africa, using species distribution modelling techniques with the sdm package in R. The result of the ensemble model, based on current climatic conditions, highlights that *Verbesina encelioides* has a high probability of occurrence in all nine provinces of South Africa, across all the projected future scenarios, namely, 2030, 2050, 2070 and 2090. Area values ranged from 810,612.09 km² in 2070, an increase of 4.23% over the current projection, to 663,356.44 km² in 2090, a decrease of 14.7% from the current projection. The outcome of these predictions showed that *V. encelioides* would benefit from the predicted climate change in South Africa. The findings could be used as a warning to implement early detection and a rapid response, or for developing one.

**Key words:** invasive plant species, golden crown-beard, Asteraceae, species distribution modelling, American dogweed

**Introduction**

Global climate change is predicted to be an important factor in the proliferation of invasive species. For instance, the increase in temperature might create new suitable habitats for some species, which might aid their establishment and thus their becoming invasive (Rahel and Olden 2008; Hellmann et al. 2008; Bradley et al. 2009; Finch et al. 2021). Therefore, in order to understand the current and future direction of population changes, it is important to incorporate the use of computational tools into predicting the effects of climate change on the availability of suitable habitats for a given species. Tools such as bioclimatic modelling are increasingly being used to project future species distributions. Species
distribution models (SDMs) are widely used tools to predict the past, current, and future habitat suitability for species using various mathematical algorithms (Adu-Acheampong et al. 2017; Shi et al. 2021; Mudereri et al. 2021; Assefa et al. 2022; Moshobane et al. 2022c). SDMs are reliant on an organismal relationship to a given geographic space and, more recently, include other factors such as land-use types (Roura-Pascual et al. 2009; Bradley et al. 2009; Pereira et al. 2010; Crossman et al. 2011; Villemant et al. 2011; Shi et al. 2021). In addition to their use in projecting current and future habitat suitability, SDMs are also useful in conservation decision-making (Pereira et al. 2010). Recently, a plethora of studies have used SDMs to predict the future distribution of invasive alien species (Urbina-Cardona et al. 2019). Recently developed modelling algorithms have suggested the use of ensemble modelling, which determines the area of agreement between different models, thereby gaining the benefits of using multiple different algorithms while better disclosing uncertainty to the end-users (Thuiller 2004; Araujo and New 2007; Marmion et al. 2009; Grenouillet et al. 2011). The ensemble approach has been used widely across different taxonomic groups—for example, insects (Zannou et al. 2022), birds (Shivambu et al. 2020a), and plants (Kaky et al. 2020; Moshobane et al. 2022b, c)—and it has consistently provided reliable results.

Verbesina encelioides (Cav.) Benth. & Hook. Fil ex Gray (Asteraceae) is a perennial herb species belonging to the family Asteraceae. It can grow up to 160 cm, with sunflower-like yellow flowers. It is a notorious weed native from North and South America – more precisely, from Mexico and from the southwestern US states of Texas and Arizona (Torrey and Gray 1969). The plant is an aggressive obligate outbreeder weed that has invaded vast expanses of pastures, orchards and forest areas in tropical and subtropical regions (Jain et al. 2008). It has extended its range within the United States to include Hawaii, to several Latin American countries, and to Australia and India (Jain et al. 2008). In India it is recorded as a perennial weed that interferes with the growth and establishment of crop species in semi-arid regions of India (Inderjit et al. 2000). The plant is regarded as naturalised in the Southwest Al-Baha region in Saudi Arabia, a hotspot area in the Arabian Peninsula (Al-Namazi et al. 2021) and in the semi-arid rangeland of southern Ethiopia (Solomon et al. 2006). It is an excellent example of an exotic invasive plant that outcompetes and extirpates native plants, and also engulfs open spaces that are used as nesting sites and disturbed habitats that appear to be ideal for native plants (Jain et al. 2008). Thus it eventually became widespread and invasive in around the world (Jain et al. 2008).

In order effectively to manage or curb the spread of *V. encelioides*, it is important to understand the potential distribution of the species. The purpose of this study was to determine the current and potential future habitat suitability for *V. encelioides*, an invasive weed in South Africa.
Materials and methods

Data-gathering

We collected data on the presence of *V. encelioides* from the Global Biodiversity Information Facility (GBIF 2022), using the rgbif package from R (Chamberlain et al. 2022), yielding a total of 500 observations. As this species is a widely distributed plant, we considered the GBIF a proper database to retrieve sufficient sampling data from a wide variety of environments where the species is present. We performed a data-cleaning routine using the CoordinateCleaner package from R (Zizka et al. 2019), deleting records with “N/A”, records in countries’ capitals and centroids, duplicate records, records with invalid or identical longitudes and latitudes, records near biodiversity institutions, and records of sightings in the sea. Afterwards, we used a 10 arc-minutes raster to delete duplicate records from within the same cell. In the end we had a total of 347 records, 25 of which were from within South Africa.

Environmental data

We downloaded current bioclimatic data from WorldClim 2.1 at the resolution of 10 arc-minutes (Fick and Hijmans 2017). As the species has a broad distribution, we excluded “quarter” variables, which often have statistical artifacts that could bias models. Then we masked variables with a one-degree-wide buffer around the presences. This guaranteed that the next step would be applied only in the environment around the records. With masked variables, we performed a multicollinearity routine using the usdm package from R (Naimi and Araújo 2016). The routine consisted in calculating the Pearson’s correlation between pairs of variables and excluding the variable with the highest variance inflation factor (VIF) within the pair with the highest correlation. This happened until the highest correlation was lower than 0.5. The remaining variables were annual mean temperature (BIO1), isothermality (BIO3), and precipitation of wettest month (BIO13). Dormann et al. (2013) suggest that 0.7 is a useful threshold in most circumstances (Dormann et al. 2013), Here the use of a low threshold, such as 0.5, returns a lower number of variables, avoiding overfitting, which is particularly problematic when aiming to model invasive species. We obtained the same set of bioclimatic variables for the new future climate change scenarios from Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al. 2016) – namely, the new Shared Socio-economic Pathway SSP585 for 2030 (mean climate between 2021 and 2040), 2050 (mean climate between 2041 and 2060), 2070 (mean climate between 2061 and 2080), and 2090 (mean climate between 2081 and 2100). Future data was downloaded for three general circulation models (BCC-CSM2-MR, IPSL-CM6A-LR, and MIROC6).
Ensemble modelling for the potential distribution of *Verbesina encelioides* in South Africa


Figure 1. The map shows the global distribution map of *Verbesina encelioides*. Red dot denotes training data used in the study, and green dot denotes test data and the projected area, South Africa, and the ensemble maps, showing the occurrence probability for *Verbesina encelioides* under current and future scenarios in South Africa. Red = high probability (1), yellow = medium probability (0.5) and light blue = low probability (0).

Species distribution modelling

For species distribution modelling, we considered a machine-learning approach, which contrasts presences and absences to obtain an occurrence probability map (Sillero and Barbosa 2021). This approach is more desirable for invasive species, since it can detect occurrences in areas with low environmental suitability. Before we began the processing, we separated South Africa’s presence data to keep it as an independent test set (25 records), with 322 records remaining for model building. As *V. encelioides* is distributed worldwide, our spatial scope was all terrestrial environments between the polar circles (Figure 1). We obtained 10 pseudo-absence datasets by randomly selecting 322 cells outside a bioclimatic envelope. The envelope
Ensemble modelling for the potential distribution of *Verbesina encelioides* in South Africa

was built with selected variables and all 347 presence records. This guaranteed that: (1) no pseudo-absence would provide the same environmental information about a presence, thus acting as a true absence; (2) there would not be any imbalance problem (Japkowicz and Stephen 2002); and (3) if any pseudo-absence dataset were biased, it would have its effects dissolved through models built with other datasets. For each set of pseudo-absences, we modelled species using five machine-learning algorithms with different approaches: classification and regression trees (CART; tree-based algorithm), support vector machine (SVM; support vector algorithm), mixture discriminant analysis (MDA; discriminant algorithm), multiple adaptive regression splines (MARS; regression algorithm); and radial basis function network (RBF; neural network algorithm). We ran algorithms four times, and used a four-fold cross-validation technique to make an inner validation, calculating the area under the receiver’s operating characteristic curve (AUC) and sensitivity for each model. In the end we had built a total of 800 models. We optimised specificity by considering the maximum sensitivity plus specificity, and deleted models with a specificity lower than 1. In invasion assessments, it is crucial to set high values of specificity to reduce omission error (type II error – i.e., the model returns an absence when the organism is in fact present). We then excluded models with a specificity lower than 1 and models with an AUC equal to 1. This second step excluded what we considered to be an overfitted model, which could have compromised our ensemble owing to lack of transferability to future climates (Vaughan and Ormerod 2005). We then calculated the mean AUC value (= 0.963) and used it as a threshold, excluding all models with an AUC value below it. We were left with 509 models, which we projected into current and future scenarios. The ensemble consisted of binarised projections, using as threshold the value that maximises sensitivity, considering the maximum sensitivity plus specificity. Afterwards, we summed the binary models into a committee average. This approach returned an uncertainty, as well as an occurrence probability map with values ranging from 0 to 1, where 1 is the value when all projections agree with a presence; 0.5 means that half of the projections agree with a presence, while the other half agree with an absence; and 0 means that all projections agree with an absence. Finally, the occupied area was obtained for each scenario, calculating the cell area and multiplying it by the predicted probability of occurrence. This returned an area value that was congruent with species occupation (Esser et al. 2019).

**Results**

The predictive ensemble model of *V. encelioides* occurrence probability is presented in Figure 1.

The result of the ensemble model, based on current climatic conditions, highlights that *V. encelioides* has a high probability of occurrence in all nine
Ensemble modelling for the potential distribution of *Verbesina encelioides* in South Africa


Figure 2. Known occurrence records for *Verbesina encelioides* in South Africa. Black dots represent GBIF records.

Table 1. Mean occupied area in square kilometers, and mean occurrence probability for *Verbesina encelioides* in South Africa. Values were obtained from ensembles for each studied scenario. SD: Standard deviation.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Mean area (km²)</th>
<th>Mean occurrence probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>777,706.40</td>
<td>0.640</td>
</tr>
<tr>
<td>SSP-585 (2030)</td>
<td>788,355.64 (SD: 7,729.38)</td>
<td>0.650 (SD: 0.01)</td>
</tr>
<tr>
<td>SSP-585 (2050)</td>
<td>799,323.45 (SD: 1,323.41)</td>
<td>0.659 (SD: 0.00)</td>
</tr>
<tr>
<td>SSP-585 (2070)</td>
<td>776,293.44 (SD: 36,373.16)</td>
<td>0.641 (SD: 0.03)</td>
</tr>
<tr>
<td>SSP-585 (2090)</td>
<td>720,423.92 (SD: 63,768.65)</td>
<td>0.596 (SD: 0.05)</td>
</tr>
</tbody>
</table>

provinces of South Africa. The central part of the country is the hotspot for *V. encelioides* invasion. Despite the suitability of Mpumalanga and KwaZulu-Natal, there are, as yet, no known records from those two provinces (Figure 2).

The mean predicted probability of occurrence and the mean area of future climatic scenarios showed an increase between 2030 and 2050, with a reduction in area predicted for between 2070 and 2090. Area values ranged from 810,612.09 km² in 2070, an increase of 4.23% in relation to current projections, to 663,356.44 km² in 2090, a decrease in 14.7% in relation to current projections. The mean occurrence probability ranged from 0.55 in 2090 to 0.67 in 2070, a decrease and an increase of 14.08% and 4.49% respectively when compared with current values. Despite the low probability of occurrence and occupied area expected for 2090, there is an increase in the probability of invasion in the Western Cape, while there is moderate decrease in suitable areas in the eastern and central parts of the country, covering mainly the Limpopo, Mpumalanga, and KwaZulu-Natal provinces (Figure 1, Table 1).
Discussion

Global climate change and human activities appear to be accelerating the spread of invasive species beyond their native geographical areas (Seebens et al. 2015, 2018; Van Kleunen et al. 2018). The present study used ensemble modelling to predict the current and potential future distribution of *V. encelioides* in South Africa under climate change. The modelling process predicted the suitable areas to be six provinces: KwaZulu-Natal, Western Cape, Eastern Cape, Limpopo, Mpumalanga, and North West. The predictions are consistent with the currently known occurrence points and records in South Africa. The probability of occurrence predicted by our ensemble model is wider than the present distribution of the species. This can be attributed to the fact that our model included only the abiotic factors – i.e., climatic variables or a lack of access to some areas; thus there are no known records from the given areas owing to private land ownership rights (Bertolino et al. 2021). Resolving this might help to identify the species that are widespread or established through a nearly comprehensive survey of all areas (Shivambu et al. 2020b).

The outcome of the future predictions showed that *V. encelioides* would benefit from the predicted climate change in South Africa, as a result of which suitable habitats would increase. Despite that, *V. encelioides* can germinate faster in low temperatures (below 20 °C) and not germinate at all in high temperatures (above 30 °C) (Karlsson et al. 2008), which explains why 2090 has lower values of probability of occurrence. Temperature increases could slow down invasion, but the cold months in South Africa present a maximum temperature lower than 20 °C for most parts of the country. Based on this, biodiversity managers and policy-makers should engage all relevant stakeholders to ensure that the species is regulated or included in the national list of invasive species, so that, where there is current demand or future need for introduction, that could be curbed (Shackleton et al. 2019; Moshobane et al. 2019, 2020b). Most importantly, there is an urgent need to control the known populations in order to stop further spreading of this invasive species.

The findings from this study are consonant with previous studies that have reported invasive species that could benefit from global climate change (Mudereri et al. 2021; Moshobane et al. 2022a). Previous studies have found that some alien species are changing their flowering season under climate change. For example, in South Africa, Scarlet Sage (*Salvia coccinea*) flowers throughout the year, while in its native range it flowers for approximately three weeks (Moshobane et al. 2020a).

Therefore, global climate change and biotic factors could enhance the spread and invasiveness of *V. encelioides*. Other factors, such as the allelopathic effects of *V. encelioides*, could also explain its success in outcompeting native species (Goel 1987; Inderjit et al. 2000): the aggressiveness and dominance of *V. encelioides* tend to overcome less aggressive native
species (Goel 1987). An in vivo study by Goel (1987) showed that *V. encelioides* extracts inhibited seed germination and reduced the growth of four species such as *Tephrosia purpurea* and *Cassia occidentalis*. As a result, it is regarded as the worst invader of peanut farms in the State of Texas (Goel 1987), competing and reducing densities and yield from agricultural products such as the peanut, *Arachis hypogaea*. (Farris et al. 2005; Farris and Murray 2006). *Verbesina encelioides* also has the potential to disrupt the annual migration of floriphilic species such as the painted-lady butterfly *Vanessa cardui*, which undertakes a multi-generational migration between Africa and Europe (Stefanescu et al. 2017).

*Verbesina encelioides* is also reported to be a favourable host of anthocorid adults and nymphs such as *Orius persequens*, *Orius tristicolor*, and *Paratriphleps laeviusculus*. Therefore, there is a potential risk to other plants through disease transmission (Calvert et al. 2019); other studies have reported more than 12 fungal pathogens that might pose a threat to native species in any alien range (Feenstra and Clements 2008). A distantly related but important study reported that *V. encelioides* induced cytotoxicity and mitochondria-mediated apoptosis in humans (Farshori 2021), unless this were harnessed for better use, such as disease treatment (Al-Oqail et al. 2016). In Bulgaria, Vassilevska-Ivanova et al. (2013) reported a hybridisation between *Helianthus annuus* and *V. encelioides*; therefore, apart from the direct impact through inhibition or competition, hybridisation might have a major impact on native species in an alien range.

**Management recommendations and future research**

On the management of *V. encelioides*, a study that evaluated post-emergence herbicides found that imazapic and imazethapyr provided sporadic control (Grichar and Sestak 1998). However, their efficacy is affected by moisture conditions and the size of the plant (Grichar and Sestak 1998). Therefore, they recommended benzaton, lactofen, pyridate, 2,4-DB, and acifluorfen or pyridate plus 2,4-DB for post-emergence control (Grichar and Sestak 1998). Alternatively, non-chemical control options could be controlled burns or watering with salt water (Feenstra and Clements 2008). Lastly, the reduction of climate change might also help to reduce the proliferation of invasive species.

Future research could attempt to predict the potential distribution with the incorporation of other variables such as anthropogenic activities, such as trading with IAS which has been shown to contribute to the proliferation of IAS (Moshobane et al. 2022). This underscores the importance of human-mediated dispersal or introduction into new areas. We strongly recommend a risk assessment and regulation of this species. Equally important is the need to develop and implement early detection and rapid response strategies.
Conclusions

To the best of our knowledge, this study is the first to model the current and future potential distribution of *Verbesina encelioides* in South Africa.

This study used ensemble modelling to predict the occurrence probability of *V. encelioides* in South Africa under current and future conditions. Our study shows that the occurrence probability in South Africa is likely to increase under future climate change. The projected current and future distributions of *V. encelioides* were mainly in the central and southeast regions of South Africa. Under the low and high emission scenarios in 2090s, there is a projected increase in suitable areas. This increase would promote the proliferation of *V. encelioides*. Although the species is not yet regulated in South Africa, clearing to reduce the spread might be a warranted action. These findings could be used as a precautionary measure for implementing early detection and a rapid response, or for developing one if none exist. Reducing the activities associated with the introduction of known and potential plant invaders might also be a critical step towards stemming the tide of invasive species. Ensemble modelling is certainly reliable in projecting the current and future distribution of species, and comes in handy in biodiversity planning and management.

Acknowledgements

The South African Department of Environment, Forestry, and Fisheries (DEFF) are thanked for funding, noting that this publication does not necessarily represent the views or opinions of DEFF or its employees. LFE was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil CAPES PhD scholarship, Finance Code 001. The study was financially supported by the Centre for Functional Biodiversity, School of Life Sciences, University of KwaZulu-Natal, and the South African National Biodiversity Institute (SANBI). We are indebted to, Louisa Mpshe who assisted with some aspects of this paper. The suggestions and comments of two anonymous reviewers and of the Handling editor Evelyn Beaury and thematic editor Catherine Jarnevich greatly improved the manuscript.

Disclaimer

Any opinion, finding, conclusion, or recommendation expressed in this material is that of the author(s), and the funding agencies do not accept any liability in this regard.

References


GBIF.org (2022) GBIF Occurrence Download, https://doi.org/10.15468/djl.pptbmn


Ensemble modelling for the potential distribution of *Verbesina encelioides* in South Africa


