Rangeland Ecology & Management 106 (2026) 108–119

Contents lists available at ScienceDirect

Rangeland Ecology & Management

journal homepage: www.elsevier.com/locate/rama

Original Research

Prediction to Prevention: Comparative Machine Learning Models for Early Intervention of *Dactyloctenium radulans* in Australia

Md Asaduzzaman [1,2,*], Mohammad Shamim Hasan Mandal [3], Eric Koetz [4], Azizur Rahman [5], Sheikh Muhammad Masum [6]

[1] *Gulbali Institute, Charles Sturt University, Wagga Wagga, NSW, Australia*
[2] *School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University, Wagga Wagga, NSW, Australia*
[3] *Japan International Research Center for Agricultural Sciences, Tsukuba, Ibaraki, Japan*
[4] *New South Wales Department of Primary Industries and Regional Development, Wagga Wagga, NSW, Australia*
[5] *School of Computing and Mathematics, Charles Sturt University, Wagga Wagga, NSW, Australia*
[6] *Department of Agronomy, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh*

ARTICLE INFO

Article history:
Received 8 December 2024
Revised 13 December 2025
Accepted 5 February 2026

Key Words:
Ecological modelling
Invasive species
Maxent
Weed management

ABSTRACT

Weeds pose a significant threat to agricultural productivity, causing economic damage. A comprehensive understanding of the distribution of weed species is essential for formulating efficient management strategies. Despite its ecological significance, limited research has been conducted on the native Australian weed species Button grass (*Dactyloctenium radulans*). In this study, four species distribution models (SDMs) were employed to estimate its potential habitat suitability across the entirety of Australia. The four SDMs used in this study were Maxent, random forest (RF), boosted regression trees, and artificial neural networks (ANN). We conducted and collected occurrence records from field surveys and also used historical occurrence records from GBIF with environmental and soil variables to build robust predictive models. Our results suggested that seasonal temperature and soil pH were critical determinants of *D. radulans* distribution, contributing up to 37% of permutation importance. Among the four SDMs, Maxent outperformed other models in predictive accuracy (area under the curve = 0.90), while RF and ANN demonstrated distinct advantages in specific contexts. The integration of field and GBIF datasets enabled us to broaden our scope, and we identified southeastern Australia as a high-risk zone for *D. radulans* proliferation. This research underscores the utility of machine learning algorithms, including Maxent in ecological modelling, providing actionable insights for weed management and habitat restoration.

Crown Copyright © 2026 Published by Elsevier Inc. on behalf of The Society for Range Management. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Introduction

Button grass (*Dactyloctenium radulans)* (R. Br.) is an Australian native grass species and an emerging cropping weed (Gradner 1952; Cameron and Storrie 2014). It can emerge in different cropping systems and competes for soil moisture and nutrients, reducing the production potential of subsequent crops (Borger et al. 2020). This species exhibits high seedbank persistence, with surface seeds remaining viable for up to 36 months and buried seeds for 42 months (Chauhan and Manalil 2022). This longevity contributes to its weedy nature, as it can rapidly replenish its seedbank. In northern Australia, *D. radulans* increases the risk of glyphosate resistance in cotton cropping system due to relay on glyphosate herbicide for weed control (Werth et al. 2011). It possesses a range of environmental stress tolerance, which further complicates its management (DPIRD 2017; Asaduzzaman et al. 2019). Further, a heavy infestation of *D. radulans* in stockyards can result in nitrate–nitrite toxicity in sheep and cattle and may also attract Australian plague locusts (Everist 1981; McKenzie et al. 2004; Sharp et al. 2019). Thus, the implementation of effective management strategies are essential to mitigate the impact of this weed.

Farmers intend to enhance profitability with modern weed management techniques, including herbicide application. Site-specific weed management (SSWM) with herbicides is an emerging approach that aims to optimize weed control while minimizing environmental impact (Shaw 2005). Currently, SSWM also

* Correspondence: Md Asaduzzaman, Building 286, School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University, Boorooma Street, Wagga Wagga, NSW 2650, Australia
E-mail address: masaduzzaman@csu.edu.au (M. Asaduzzaman).

https://doi.org/10.1016/j.rama.2026.02.004
1550-7424/Crown Copyright © 2026 Published by Elsevier Inc. on behalf of The Society for Range Management. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

integrates weed-sensing systems, decision models, and precision implements to target specific weed-infested areas (Christensen et al. 2009; Ghatrehsamani et al. 2023; Sarkar et al. 2024). Further advancements in technologies, including 3D cameras, multispectral imaging, followed by machine learning approach, have improved weed recognition and classification capabilities, making SSWM approach more economically feasible for farmers (Shorewala et al. 2021; Gerhards et al. 2022). However, challenges remain, including the need for accurate prediction for potential future infestation in suspected hot spots (Christensen et al. 2009), which requires the integration of robust datasets, advanced modelling techniques, and consideration of dynamic environmental and biological factors influencing weeds' infestation trends.

Species distribution models (SDMs) are popular tools to evaluate both the current and potential distribution of targeted species by analyzing environmental variables alongside spatial maps (Hengl et al. 2009; Rios et al. 2024). These models aid in identifying and assessing the influence of various ecological factors on weed distribution, enabling the prediction of species occurrence (Franklin 2013). A robust predictive model supports proactive management by identifying high-risk areas for weed invasion. SDMs enable researchers to prioritize control measures, thereby optimizing resource allocation (e.g., SSWM) and minimizing environmental impact (Bradley et al. 2010; Jiménez-Valverde et al. 2011). SDM outcomes can further guide habitat restoration by identifying areas for re-establishing native vegetation, enhancing natural resistance to weed establishment, and spread (Thuiller et al. 2005; Peterson et al. 2011). While most SDMs rely on existing data to avoid the high costs associated with extensive survey-based sampling (Powney and Isaac 2015; Johnson et al. 2020), these models often fail to fully capture the complexity of biotic interactions, such as competition and predation, which can be essential for accurate ecological predictions. Furthermore, the reliability of these models is heavily influenced by the quality of historical data, which can be biased or inaccurate, thereby compromising the accuracy of model outputs (Guisan and Thuiller 2005; Peterman et al. 2013; Fourcade et al. 2014). Addressing these limitations requires incorporating detailed ecological data and refining models with traditional habitat survey information. This integration enhances the reliability of weed distribution projections and supports improved decision-making in weed management.

Machine learning models such as random forest (RF), boosted regression trees (BRT), Maximum Entropy Modelling (Maxent) and artificial neural networks (ANNs) have become widely used for predicting species' habitat suitability due to their flexibility, robustness and ability to manage and analysis complex high dimensional ecological data (Phillips et al. 2006; Franklin 2010; Elith et al. 2011). Each offer unique strength and different level of predictive accuracy. For example, RF can accommodate a large number of predictor variables and capture nonlinear relationships, which are common in ecological datasets (Cutler et al. 2007; Evans et al. 2011). BRT effectively captures complex nonlinear relationships by building decision tree, and particularly useful for identifying key environmental factors in species distribution (Elith et al. 2008). Maxent, ideal for presence-only data, estimates habitat suitability by maximizing entropy, providing clear maps and threshold insights for conservation in data-limited areas (Phillips et al. 2006; Phillips and Dudík 2008). ANN, though less transparent, excel in handling large, interactive datasets, making them suitable for complex ecological modelling (Olden et al. 2008). It is generally recommended to test a range of modelling techniques, as no single method is universally superior in all contexts (Johnson and Gillingham 2004; Hirzel and Le Lay 2008). Adopting multiple approaches can enhance predictive accuracy and provide more robust insights into weed distribution patterns. However, most studies on weed SDMs focus on individual modelling approaches and provide conclusions based on a single method. Previous distribution mapping of *D. radulans* has primarily relied on environmental variables. However, the current understanding of its distribution patterns, particularly in relation to the interactions among environmental factors, soil properties, and land-use patterns across Australia, remains limited. This study presents several novel contributions to weed ecology and species distribution modelling. This is the first comprehensive assessment integrating four distinct machine learning algorithms (Maxent, RF, BRT, and ANN) with a combined dataset of field surveys and historical records. Second, we incorporated soil properties alongside climatic variables, addressing a critical gap in understanding the edaphic requirements of *D. radulans* distribution. Third, our comparative modelling framework provides methodological insights into algorithm performance for weed SDMs, contributing to best practices in SSWM. This multifaceted approach enables more robust predictions and provides a validated framework for modelling this emerging weed species in Australian cropping systems and to addresses these knowledge gaps by investigating three key research questions:

(1) Which environmental and soil parameters are crucial determinants of *D. radulans* habitat suitability?
(2) Which modelling approach yields superior predictive accuracy in generating habitat suitability maps?
(3) Based on the optimal predictive model, which geographical areas should be prioritized for *D. radulans* management?

To achieve these objectives, this research integrates both physical survey data and historical records to evaluate multiple modelling techniques, thereby providing a comprehensive assessment of predictive accuracy.

Methods

Collection of occurrence records

Field surveys to record the occurrence and absence of *D. radulans* were conducted in November 2019, focusing on cotton-growing regions in northern New South Wales (NSW), Australia. The study area includes agricultural landscapes characterized by cotton production and related land-use practices. A systematic sampling approach was employed to ensure representative coverage of the area (Asaduzzaman et al. 2018). GPS coordinates for the 47 sampling sites were recorded using a handheld GPS device (Garmin GPSMAP 67 series) (Figs. 1 and 2). Among the sampling sites, 22 were recorded as presence locations and 25 as true absence locations. Images of the species occurring in the landscapes were also collected. All field data collection and sampling protocols adhered to ethical guidelines and permissions from relevant authorities and landowners/managers. Data sharing and confidentiality were maintained throughout the study to protect sensitive information and comply with privacy regulations. In addition to our existing field survey, we collected occurrence records of *D. radulans* from the GBIF (gbif.org) database using the "spocc" package in R (Chamberlain 2021). Erroneous and duplicate points within each 1 km^2 raster grid were removed from the original dataset. After combining with field-surveyed occurrence points, we had 187 locations for further analyses.

Environmental and bioclimatic data

To depict current climate conditions, we used 19 bioclimatic variables from the WorldClim database (worldclim.org) (Table 1), spanning almost 45 years (1970–2015) (Fick and Hijmans 2017). The datasets spatial resolution was approximately 1 km^2. We also obtained a digital elevation model (about 1 km^2) from the WorldClim database and computed the slope using the "raster" package

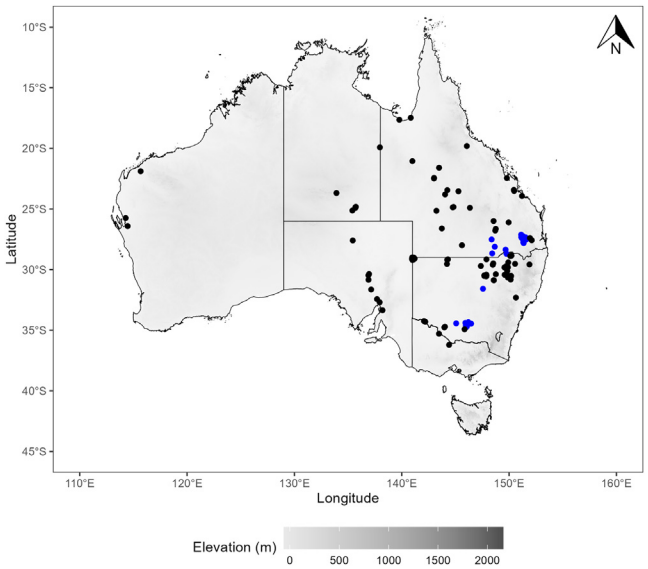


Figure 1. Spatial distribution of *Dactyloctenium radulans* occurrence records: Field survey data (blue) and GBIF.org records (black) overlayed on mean elevation from sea level (WorldClim.org) for Cotton-Growing Regions of Australia. The elevation data is shown as a gradient (indicating elevation levels from sea level to higher altitudes), which can provide context for understanding how *D. radulans* distribution correlates with topographic features.



Figure 2. Field survey of *Dactyloctenium radulans* infestation in cotton fields. Left: Cotton field with *D. radulans* co-occurring with other weed species. Right: Cotton field heavily infested with *D. radulans*, indicating severe competition for resources.

for R (Hijmans 2020). The land-use data is vital for modelling and forecasting the habitats of *D. radulans* as it is found in both agricultural and nonagricultural regions of Australia (Asaduzzaman et al. 2018). We obtained 20-year (2001–2020) average land cover of the Terra and Aqua combined Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) Version 6 data (Friedl and Sulla-Menashe 2019) using the Google Earth Engine platform (Gorelick et al. 2017). We selected the classification scheme of the Annual International Geosphere-Biosphere Programme (IGBP), which includes 17 distinct land-use categories such as forests and croplands (Friedl and Sulla-Menashe 2019), which were aggregated to five common land cover types. The data, with a 500 m^2 resolution, was transformed to WGS84 (EPSG: 4326) and aligned with bioclimatic variables through resampling (Friedl and Sulla-Meanashe 2022). The land-use data is crucial for modelling and forecasting the habitats of *D. radulans*, as the species is found across both agricultural and nonagricultural regions of Australia (Asaduzzaman et al. 2018). We also collected nine soil attributes (Table 1) representing the average values of top 60 cm soil from the Soil and Landscape Grid National Soil Attribute Maps by CSIRO (Rossel et al. 2015) from Google Earth Data Catalog (Gorelick et al. 2017). All the data were rescaled to approximately 1 km^2 to match the bioclimatic layers and were used as predictors for subsequent model building.

Table 1
Descriptions of bioclimatic, topographic, and soil variables selected for modelling habitat suitability of *Dactyloctenium radulans* in Australia.

SL	Code	Name	Unit
1	bio1	Annual mean temperature	°C
2	bio2	Mean diurnal range (mean of monthly [max temp–in temp])	°C
3	bio3	Isothermality (Bio2/Bio7) (×100)	
4	bio4	Temperature seasonality (standard deviation × 100)	°C
5	bio5	Max temperature of warmest month	°C
6	bio6	Min temperature of coldest month	°C
7	bio7	Temperature annual range (BIO5–BIO6)	°C
8	bio8	Mean temperature of wettest quarter	°C
9	bio9	Mean temperature of driest quarter	°C
10	bio10	Mean temperature of warmest quarter	°C
11	bio11	Mean temperature of coldest quarter	°C
12	bio12	Annual precipitation	mm
13	bio13	Precipitation of wettest month	mm
14	bio14	Precipitation of driest month	mm
15	bio15	Precipitation seasonality (coefficient of variation)	1
16	bio16	Precipitation of wettest quarter	mm
17	bio17	Precipitation of driest quarte	mm
18	bio18	Precipitation of warmest quarter	mm
19	bio19	Precipitation of coldest quarter	mm
20	soil1	pH	-
21	soil2	Sand	%
22	soil3	Silt	%
23	soil4	Clay	%
24	soil5	Soil organic carbon	%
25	soil6	Available water capacity	%
26	soil7	Effective cation exchange capacity	meq/100 g
27	soil8	Total nitrogen	%
28	soil9	Total phosphorus	%
29	elv	Elevation	m
30	lc	Land use category	-

Addressing sampling bias and spatial autocorrelation

To address potential sampling bias in GBIF data (Beck et al. 2014; Fourcade et al. 2014), we applied spatial thinning by removing duplicate records within each 1 km^2 raster cell, reducing overrepresentation from well-sampled areas (Aiello-Lammens et al. 2015). For background point selection in Maxent, we drew 10 000 background points at random from locations where other grass species were recorded in GBIF. This approach reduces sampling bias by aligning background selection with areas that share similar collection effort. However, because model evaluation used random partitioning rather than spatially explicit cross-validation, residual spatial autocorrelation may influence performance estimates. We acknowledge this as a limitation, particularly in historically undersampled regions such as Western Australia. Another technical reason for not using spatial blocking techniques, such as the checkerboard method or blockCV, is that our occurrence records were relatively sparse. Applying these methods would result in very few records per block, making model training unstable and complicating fair comparisons among algorithms like Maxent, RF, BRT, and ANN.

Modelling techniques

We selected Maxent, RF, BRT, and ANN based on their established performance in species distribution modelling and distinct algorithmic strengths (Elith et al. 2006; Cutler et al. 2007; De'ath 2007; Olden et al. 2008). Maxent was chosen for its proven efficacy with presence-only data and widespread application in ecological studies (Phillips et al. 2006). RF and BRT were selected for their ability to handle complex nonlinear relationships and variable interactions common in ecological datasets (Elith et al. 2008). ANN was included to capture potential high-order

interactions that tree-based methods might miss (Lek and Guégan 1999). While ensemble approaches and algorithms such as Support Vector Machines, XGBoost have demonstrated utility in species distribution modelling (Thuiller et al. 2009; Valavi et al. 2019), we did not include them because our primary objective was to assess algorithm-specific performance and interpretability for this understudied weed species. The four selected models represent distinct methodological paradigms: presence-only, tree-based, boosting, and neural networks. This diversity enables a comprehensive comparison. Our approach enables us to 1) evaluate individual algorithm strengths and weaknesses, 2) provide transparent insights into variable importance across different modelling frameworks, and 3) establish baseline performance metrics for future ensemble applications. Furthermore, recent meta-analyses suggest that well-tuned individual models often perform comparably to ensemble predictions, particularly when data quality is high (Hao et al. 2019). Therefore, our strategy allows identification of the single best-performing model, which is computationally efficient for operational management applications where rapid updates are required (Valavi et al. 2019).

Maxent modelling approach

The Maxent algorithm requires background locations where the species were not present, and we randomly choose 10 000 points as background points. The data was split into an 80% training set and a 20% testing set. We fine-tuned the model using cross-validation, a grid search method, and the Maxent algorithm from the SDMtune package (Vignali et al. 2020). We evaluated a model with different combinations of regularization multiplier (RM) values and feature class (FC) combinations. RM values were used to smooth the model, penalizing more complex models with higher RM values and resulting in simpler models (Elith et al. 2011). We tested seven feature class combinations, viz., L, LQ, LH, LQP, LQPH, and LQHPT. L represents linear, Q represents quadratic, H represents hinge, P represents product, and T represents threshold. Therefore, constructing one model for each of 19 RM values (0.2–2.0, increment of 0.1) and 6 FC combination resulted in 114 models. We utilized the area under the curve (AUC) values for model validation (Radosavljevic and Anderson 2014; Kass et al. 2021).

To address multicollinearity among predictor variables, we applied a data-driven variable selection procedure implemented in SDMtune package (Vignali et al. 2020), which iteratively removes highly correlated variables (Spearman $|\rho| > 0.7$) based on permutation importance and AUC. This approach is widely used in single-species SDM studies where interpretability of original predictors is critical (Dormann et al. 2013). We made further adjustments to exclude less important variables (permutation importance of less than 5) based on the final model settings (RM: 1; FC: LQPH). The final model's enactment was estimated using the completely concealed test data set.

We used the final variables selected in Maxent, which already addressed multicollinearity issues. This approach ensures fair and stable comparisons among algorithms and prevents instability that could arise from using different datasets.

Random forest

One of the most accurate models for classification or regression trees is RF, which creates several decor-related regression trees using a bagging technique after selecting numerous subsamples from the data using bootstrap aggregation (Breiman 2001). To ensure that every tree in the forest is dependent on randomly selected values, RF combines tree predictors to provide a comparable distribution for every tree in the forest (Breiman 2001). We used RF to create an ensemble of unpruned classification or regression trees using bootstrap samples and randomly selected features during tree induction. We set Mtry to 2, optimizing the balance between randomness and predictive power. The RF model consisted of 400 decision trees, each contributing to the overall prediction based on a subset of the training data. The trained RF model was evaluated using the test dataset to assess its performance in predicting weed presence and absence across different environmental and soil conditions. The performance measurements used included accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC-ROC).

Boosted regression trees

The combined dataset was partitioned into training and testing subsets, preserving the distribution of *D. radulans'* presence and absence instances. For training the BRT model, the model consisted of 700 regression trees, each contributing to the overall prediction based on a subset of the training data. We limited the interaction depth to 1, controlling the complexity of the model and preventing overfitting. The shrinkage parameter was set to 0.1, controlling the contribution of each tree to the final prediction and improving model stability. We used a bag fraction of 0.5, indicating the proportion of training data randomly sampled at each iteration to build individual trees. The trained BRT model was evaluated using the test dataset to assess its performance in predicting weed presence and absence across different environmental and soil conditions. The model's predictive performance was assessed using accuracy, sensitivity, specificity, and AUC-ROC. In the BRT model, we used the "dismo" package v1.3-14 (Ridgeway 2006) with an additional recommended configuration code by Elith et al. (2008).

Artificial neural network

The unified dataset was divided into training and testing subsets, preserving the distribution of *D. radulans'* presence and absence instances. The decay parameter, set to 0.1, controls the rate at which weights in the network are adjusted during training, preventing overfitting and improving generalization. The range parameter, set to 0.7, determines the range of initial weight values, influencing the model's learning dynamics and convergence speed. The maximum number of iterations (maxit) was set to 300, defining the maximum number of training epochs or iterations during model training. The trained ANN model underwent rigorous validation using the test dataset to accurately assess its ability to predict weed presence. The model's effectiveness and robustness were assessed using performance metrics including accuracy, sensitivity, specificity, and AUC-ROC.

Habitat suitability map of D. radulans

After evaluating the models, the best model was used to estimate the probability map of habitat suitability. This map ranging from 0 to 1, representing the probability of species habitat suitability, with 0 indicating the lowest probability and 1 indicating the highest probability of species occurrence. For interpreting the map more clearly this map was further categorized into five habitat suitability groups: 1) not appropriate ($\leq$0.20), 2) least suited (0.21–0.40), 3) moderate (0.41–0.60), 4) high (0.61–0.80), and 5) extremely high ($>$0.80) (Qin et al. 2017; Abdelaal et al. 2019; Mandal et al. 2023). The areas for each suitability class were calculated in square kilometers, along with the corresponding percentage changes.

Results

Variable importance under current scenario

The variable importance analysis provided key insights into the primary factors influencing weed distribution, highlighting the

Table 2
Variable permutation importance in different models.

Variable	ANN		BRT		Maxent		RF	
	Importance	Sd	Importance	Sd	Importance	Sd	Importance	Sd
bio2	14.6	0.002	11.2	0.001	8.9	0.003	4.1	0.000
bio4	32.1	0.001	30.7	0.007	37.1	0.013	34.8	0.003
bio14	-	-	-	-	8.1	0.00	-	-
bio18	33.7	0.002	14.7	0.004	15.9	0.004	6.3	0.001
soil1 (pH)	19.7	0.000	31.6	0.002	30.0	0.001	33.4	0.001
soil8	0.0	0.000	11.9	0.002	-	-	21.4	0.001

Table 3
Threshold evaluation and performance metrics for Maxent, RF, BRT, and ANN.

Model	Threshold	Cloglog value	Fractional predicted area	Training omission rate	Test omission rate	*P* values
ANN	Minimum training presence	0.000	0.975	0.000	0.000	0.000
	Equal training sensitivity and specificity	0.008	0.217	0.001	0.000	0.000
	Maximum training sensitivity plus specificity	0.028	0.087	0.002	0.000	0.000
	Equal test sensitivity and specificity	0.007	0.243	0.001	0.000	0.000
	Maximum test sensitivity plus specificity	0.007	0.241	0.001	0.000	0.000
BRT	Minimum training presence	0.000	0.995	0.000	0.000	0.000
	Equal training sensitivity and specificity	0.018	0.109	0.001	0.000	0.000
	Maximum training sensitivity plus specificity	0.017	0.114	0.001	0.000	0.000
	Equal test sensitivity and specificity	0.010	0.171	0.000	0.000	0.000
	Maximum test sensitivity plus specificity	0.009	0.183	0.000	0.000	0.000
Maxent	Minimum training presence	0.005	0.802	0.000	0.000	0.000
	Equal training sensitivity and specificity	0.197	0.153	0.001	0.000	0.000
	Maximum training sensitivity plus specificity	0.168	0.176	0.001	0.000	0.000
	Equal test sensitivity and specificity	0.151	0.192	0.001	0.000	0.000
	Maximum test sensitivity plus specificity	0.142	0.202	0.110	0.000	0.000
RF	Minimum training presence	0.592	0.000	0.002	0.000	0.000
	Equal training sensitivity and specificity	0.593	0.000	0.002	0.000	0.000
	Maximum training sensitivity plus specificity	0.593	0.000	0.002	0.000	0.000
	Equal test sensitivity and specificity	0.010	0.091	0.000	0.000	0.000
	Maximum test sensitivity plus specificity	0.013	0.072	0.000	0.000	0.000

most significant predictors for habitat suitability (Table 2). For instance, seasonal temperature (bio4) and soil pH (soil1) were identified as crucial factors, contributing 37% and 30%, respectively, to the overall permutation importance. Both RF and BRT models further validated the importance of these variables in predicting habitat suitability for *D. radulans*. In contrast, ANN model indicated that precipitation during the warmest quarter (bio18) was a more dominant predictor than soil pH, accounting for approximately 34% of the permutation importance. Additionally, other environmental variables, such as the mean diurnal range (bio2), were consistently recognized as important, contributing 4% to 15% of the permutation importance across all four models, depending on the model type.

The jackknife test was used to further evaluate the importance of various predictor variables (Fig. 3). When tested individually, among variables, bio4 and soil1 yielded the highest for all four models. The models' resilience to the exclusion of each variable suggests that it is not overly reliant on any single predictor. This redundancy indicates that these variables are robust, with a flexible structure that allows for accurate predictions even if certain predictors are unavailable.

Thresholds, predictive capabilities, and model reliability

Threshold values obtained from four models were carefully selected to balance sensitivity and specificity in predictions. The comparative analysis of the Maxent, RF, BRT, and ANN models reveals distinct patterns in cloglog values, fractional predicted areas, omission rates, and statistical significance across varying threshold criteria (Table 3). All models demonstrated statistical significance and zero omission rates across thresholds, indicating robust predictive accuracy. Maxent showed the most balanced performance, with consistent cloglog values (0.005–0.197) and low fractional predicted areas (0.001 for most thresholds, rising only to 0.110 at the maximum test sensitivity plus specificity), demonstrating adaptability across sensitivity and specificity requirements. In comparison, RF produced higher cloglog values (above 0.592) for training-oriented thresholds and lower values (0.010–0.013) for test-oriented thresholds, with a stable fractional predicted area of 0.002, dropping to 0.000 for certain test thresholds. BRT exhibited the widest range in cloglog values (from 4.32e – 05 to 0.1833), with most thresholds maintaining a low predicted area (0.001), though the minimum training presence reached a high of 0.995, suggesting greater sensitivity to threshold adjustments. ANN also performed consistently, with cloglog values from 7.98e – 05 to 0.243 and fractional predicted areas between 0.000 and 0.002. Overall, Maxent stands out for its balance and flexibility, while RF and ANN deliver stable but less adaptable performance. BRT's broader threshold sensitivity makes it suitable for applications needing extended spatial predictions under minimal presence criteria.

Model performance and AUC ratio

We evaluated single SDMs like Maxent, RF, BRT, and ANN demonstrated good predictive abilities with mean AUC values ranging from 0.92 to 0.95 and accuracy scores within acceptable ranges (Fig. 4). The RF and ANN showed sensitivity to iterations, with RF frequently demonstrating better predictive performance. The modelling method impacted predicted potential distributions for an independent area. The AUC yielded consistent outcomes across all models. The AUC values for *D. radulans,* for instance, are approximately 0.90 despite the result clearly demonstrating a difference. The results indicated that the average AUC values for *D. radulans* were all greater than 0.90 for the four models. Even with a high AUC value, not every model built using the training dataset could accurately predict the presence of the studied species in some known locations where it exists (Fig. 4). For example, the

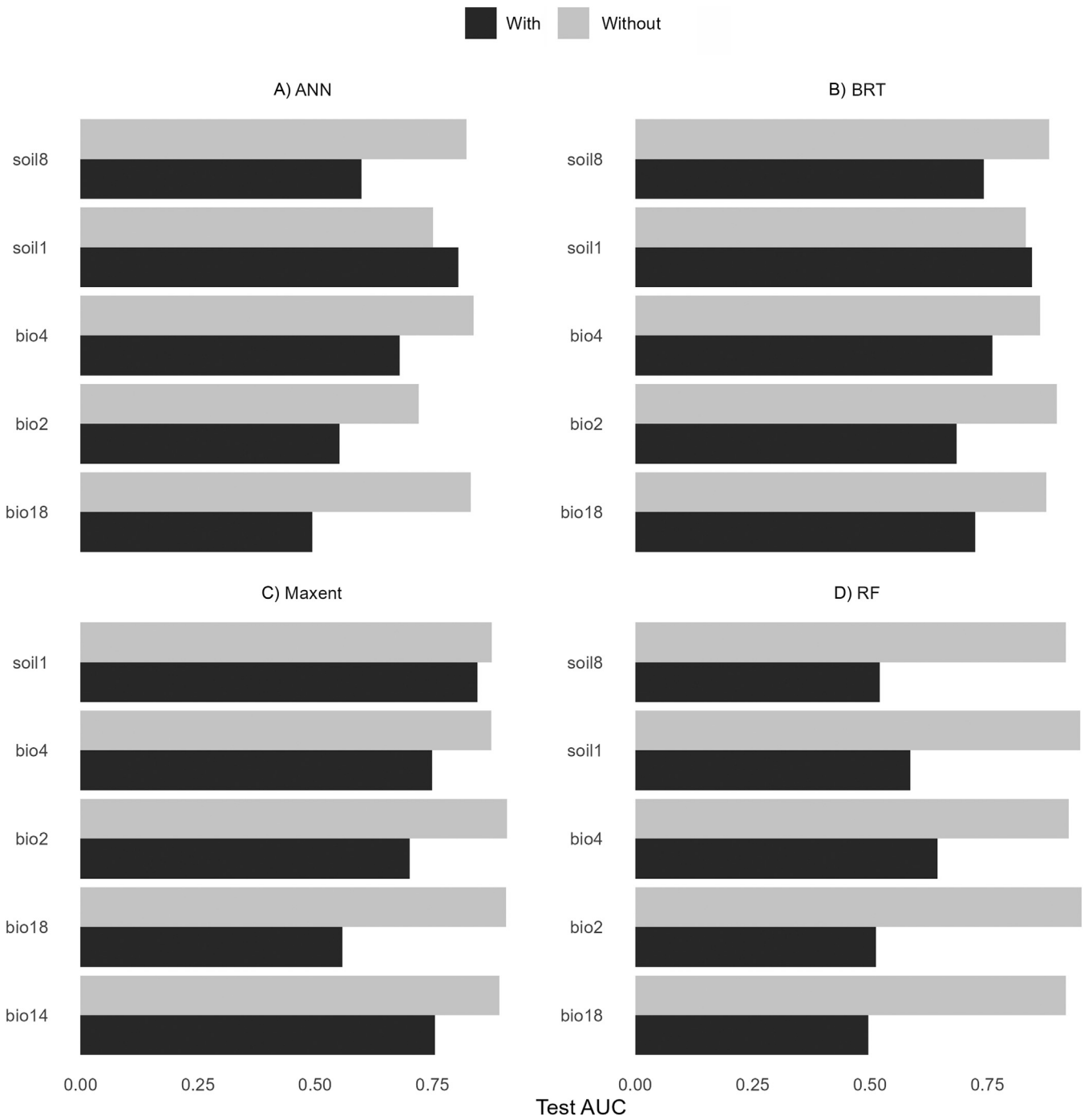


Figure 3. The jackknife test evaluated the importance of various predictor variables. The performance of each variable was assessed under two conditions: using each predictor in isolation (with) and excluding each predictor from the model (without).

AUC for *D. radulans* in the RF and BRT models was 0.90 and 0.92, respectively, while only 0.90 in the Maxent model.

Comparison of habitat suitability predictions

The Maxent model predicted the largest area of suitable habitat, particularly in southeastern Australia, with smaller areas in the east and northeast (Fig. 5). It identified broader suitable regions, especially in higher-elevation, ecologically complex areas. In contrast, the BRT model offered a more conservative prediction, with few areas reaching moderate suitability and limited to small patches in eastern Australia. RF provided a more moderate estimate, identifying a few moderately suitable patches in eastern and northern Australia, offering a balance between the expansive Maxent model and the restrictive BRT model. ANN showed a fragmented distribution of suitable areas, with patterns similar to RF.

Figure 6 illustrates the habitat suitability of *D. radulans* across Australia based on Maxent modelling, with detailing the distribution of suitability classes (Appendix A1). The majority of the landscape (6 549 017 km^2 or 78.5%) is classified as not suitable, primarily covering the arid regions of central and western Australia, where low precipitation and poor soil fertility inhibit the weed's establishment. In contrast, approximately 1 036 529 km^2 (12.4%) is deemed suitable, with varying degrees of suitability. Least suitable

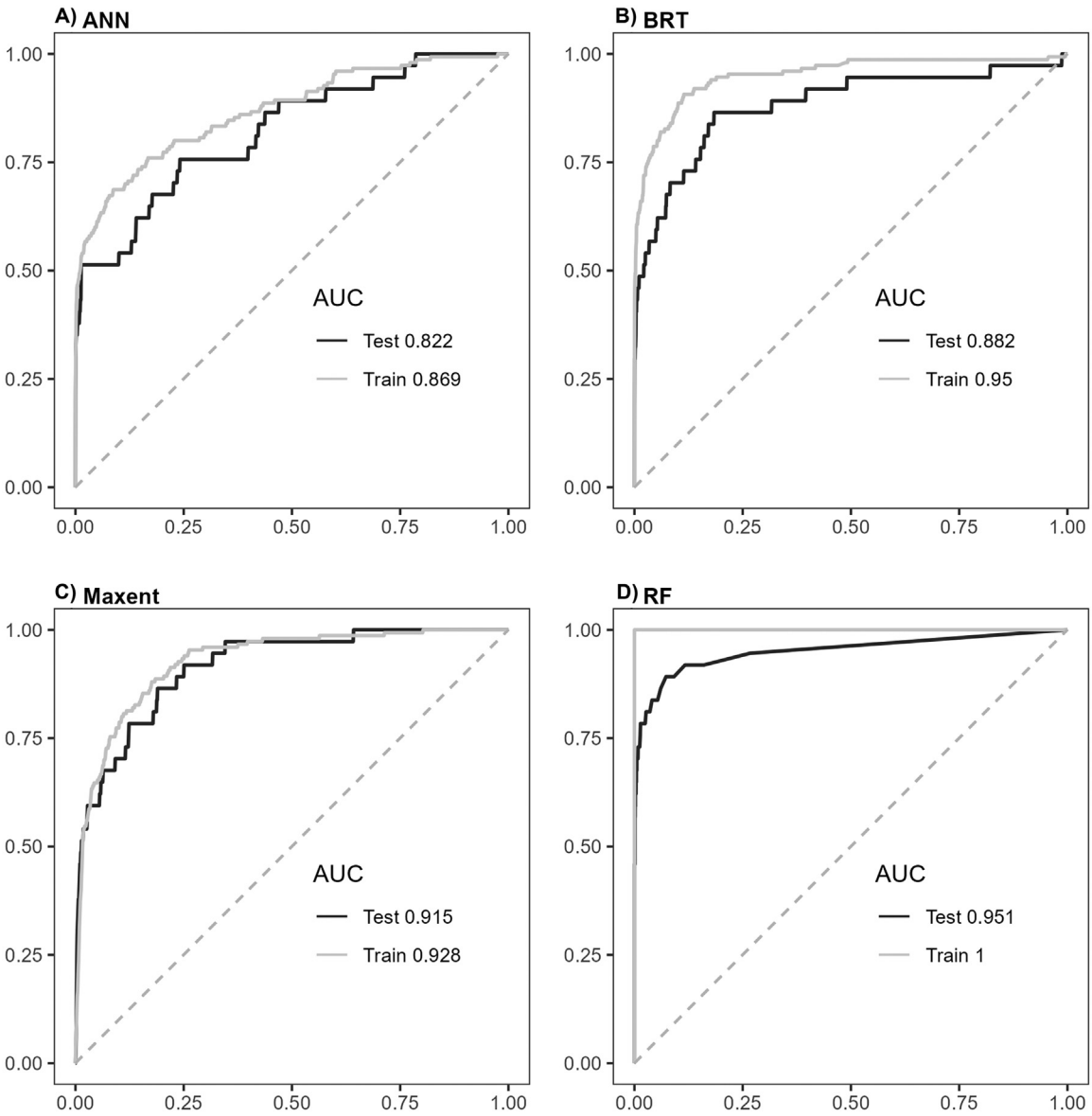


Figure 4. The receiver operating characteristic (ROC) curve for **(A)** ANN, **(B)** BRT, **(C)** Maxent, and **(D)** RF, with the *y*-axis representing sensitivity and the *x*-axis representing the false positive rate (1−specificity). The straight dotted line indicates the behavior of a random classifier.

areas (543 888 km^2 or 6.5%) are in transitional zones between arid and semiarid regions, while moderate suitability (237 550 km^2 or 2.9%) is concentrated in the eastern coastal plains and inland areas with moderate rainfall. High suitability (131 328 km^2 or 1.6%) is found in agriculturally significant regions such as the Murray–Darling Basin and NSW, whereas very high suitability (123 763 km^2 or 1.5%) is concentrated in southeastern Australia, where optimal climatic and soil conditions support the weed's growth.

Discussion

Projecting the distribution of species richness is a challenging task, and SDMs provide an initial representation of species ranges that require continuous updates as new data becomes available or environmental conditions shift (Guisan and Thuiller 2005). Models based solely on environmental variables may overestimate species ranges, as they fail to consider factors such as historical absences, dispersal barriers, soil, or biotic interactions (Pulliam 2000; Jiménez-Valverde et al. 2008). Therefore, creating an accurate demonstration of species richness requires a thorough analysis to identify areas of omission and incorporate presences in locations where predictions did not anticipate their occurrence (Heikkinen et al. 2014). We conducted an in-depth process of model refinement and selection to identify the most effective predictive models for weed distribution, focusing on four machine learning techniques. The models were optimized through careful parameter adjustments and variable selection to improve their robustness against collinearity between predictors, which could otherwise distort the results. Our collinearity analysis showed no significant correlations among the chosen variables, confirming the models' validity and reliability. The high AUC values across all models reflect their ability to distinguish suitable habitats for weed species from unsuitable ones, despite slight differences in performance. For example, while Maxent and RF models exhibited comparable AUC values for BRT and ANN models appeared more sensitive to environmental variations, which aligns with findings from past research indicating that these algorithms excel in modelling complex ecological phenomena (Elith et al. 2008). Nevertheless, it is essential to note that high AUC values alone do not guarantee accurate predictions in real-world conditions, as

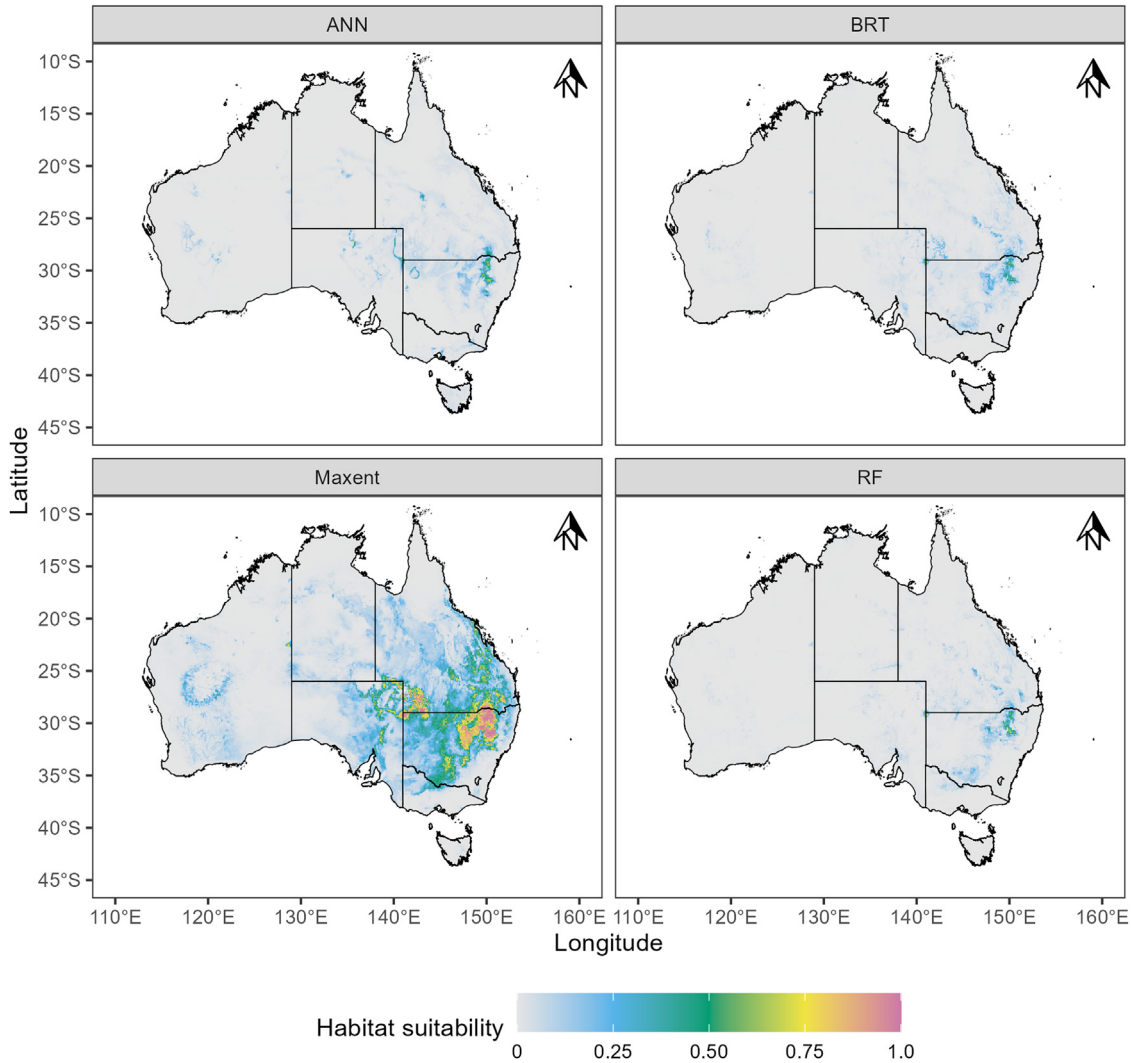


Figure 5. Comparative habitat suitability predictions for *Dactyloctenium radulans* across Australia using four machine learning models: **(A)** Maxent, **(B)** RF, **(C)** BRT, and **(D)** ANN. State boundaries are shown for reference. Coordinate system: WGS84 (EPSG: 4326). Note differences in spatial extent of predicted suitable areas, with Maxent identifying the broadest distribution, particularly in southeastern regions.

evidenced by some models' inability to predict weed presence in known locations (Lobo et al. 2008; Peterson et al. 2011).

The variable importance revealed that climatic and edaphic factors significantly influence the distribution of *D. radulans*. Across all models, seasonal temperature (bio4) and soil pH (soil1) were consistently identified as critical predictors, contributing between 30% and 37% of the permutation importance. These results align with previous studies emphasizing the pivotal role of temperature and soil conditions in determining species distributions (Peterson et al. 2011). The results further align with Asaduzzaman et al. (2019), who demonstrated that temperature is a key factor in germination and emergence of *D. radulans* and that this species can withstand various soil pH levels (Asaduzzaman et al. 2019). Interestingly, precipitation of the warmest quarter (bio18) emerged as the most influential variable in the ANN model, highlighting the model's sensitivity to precipitation-related parameters. The differences in variable importance across models reflect the unique strengths of each algorithm in capturing species-environment relationships. For instance, Maxent and RF placed greater emphasis on temperature and soil variables, ANN demonstrated higher sensitivity to precipitation, suggesting that neural networks may be better suited for capturing complex nonlinear relationships between environmental variables and *D. radulans* distributions. This variation across models reinforces the importance of using multiple modelling techniques to account for the diverse range of ecological factors influencing species distribution. A key aspect of model performance was the evaluation of threshold values to balance sensitivity and specificity. The BRT exhibited a wider range of threshold sensitivities, which can be useful for applications requiring minimal presence criteria but may introduce noise in areas with low species occurrence. The species response curve to different variables used for modelling reveals the relationship between environmental variables and suitable habitat for *D. radulans*. This highlights the potential of using multiple bioclimatic and soil variables in a complementary manner to effectively predict *D. radulans* distribution. The insights gained from this analysis can inform future studies aiming to optimize ecological niche models, particularly in contexts where predictor data may be limited or uncertain. Further, the Maxent model showed consistent cloglog values and low fractional predicted areas, making it the most adaptable to different threshold settings. Earlier research showed that Maxent has consistently demonstrated its effectiveness as a modelling tool (Elith et al. 2006; Graham and Hijmans 2006; Hernandez et al. 2006). It is particularly notable for its perfor-

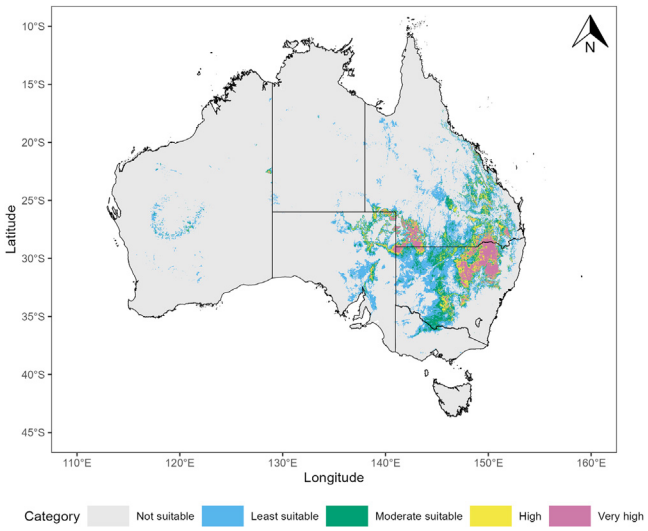


Figure 6. Predicted habitat suitability of *Dactyloctenium radulans* across Australia based on the best-performing Maxent model. Northern NSW and North-Western NSW (yellow to mauve), and particularly the Murray–Darling Basin-based cotton-growing regions of New South Wales, show the highest habitat suitability.

mance in dealing with spatial uncertainty in species occurrence data, even when there are as few as four records. This makes Maxent a valuable tool for modelling a wide range of species, especially when dealing with limited or imprecise location data.

The integration of field survey data with GBIF data allowed for a comprehensive assessment of weed distribution across different spatial and temporal scales. The overlap between occurrence points from both datasets enhanced confidence in the reliability of the predicted distributions. This integration highlights the value of combining datasets with complementary strengths field survey data provide detailed, localized insights, while GBIF data offer broad spatial and temporal coverage. Such approaches are increasingly recognized as essential for robust ecological assessments, particularly in the context of invasive species management (Beck et al. 2022). However, discrepancies between the datasets also revealed potential sources of bias and uncertainty. Areas with only red (field survey) or blue (GBIF) dots could indicate sampling biases, differences in data collection methods, or variations in *D. radulans*' habitat preferences. Investigating these discrepancies is crucial for refining predictive models and improving data reliability. Understanding these discrepancies provides a deeper insight into the limitations and strengths of each dataset (Lobo et al. 2008). Integrating field surveys with data from the GBIF can significantly enhance the understanding of weed distributions. This combination not only validates and refines predictive models but also helps identify priority areas for management actions and improves the overall reliability of ecological data (Tingley et al. 2012). By merging datasets, researchers can leverage the strengths of each and overcome individual limitations, leading to more robust ecological assessments (Peterson et al. 2011). Furthermore, validating GBIF data against field surveys using expert knowledge can enhance confidence in the model outcomes, as highlighted in previous ecological studies (Yoccoz et al. 2001). Future research should focus on developing methods to harmonize disparate datasets and address temporal inconsistencies, thereby enhancing the reliability of ecological predictions (Araújo and Guisan 2006).

The habitat suitability analysis identifies southeastern Australia as a critical hotspot for the potential spread of *D. radulans*, driven by favorable conditions such as high precipitation, moderate temperatures, and fertile soils. The Murray–Darling Basin emerges as a particularly vulnerable region, where the weed could pose a significant threat to native ecosystems and agricultural productivity. Additionally, fragmented zones of moderate to high suitability along the eastern Queensland coast indicate that *D. radulans* may capitalize on localized environmental factors, such as seasonal rainfall and soil pH, necessitating targeted monitoring to prevent future invasions. Although 78.5% of Australia is classified as unsuitable for the weed, its ability to establish in specific high-risk regions highlights the need for early detection and localized management. This distribution aligns with prior research highlighting the species' preference for semiarid to moderately wet regions with fertile soils and moderate elevations. Additionally, fragmented zones of moderate to high suitability were observed along the eastern Queensland coast, influenced by localized factors such as precipitation and soil pH. The high predictive accuracy of the Maxent model (AUC = 0.90) underscores its effectiveness in identifying potential invasion zones, guiding weed management efforts. Future research should integrate climate change scenarios to refine distribution predictions and enhance control measures through real-time environmental monitoring, particularly in vulnerable agricultural areas.

Management implications and policy recommendations

The identification of southeastern Australia and the Murray–Darling Basin as high-risk zones (Fig. 6) indicates the necessity for targeted programs and early detection systems in these regions. We recommend establishing a tiered management approach: 1) intensive monitoring in very high suitability zones covering 123 763 km^2, particularly in cotton-growing regions of NSW; 2) preventive measures in high and moderate suitability areas (368 878 km^2) to prevent establishment, and 3) exclusion protocols to prevent introduction into currently unsuitable areas that may become vulnerable under changing climatic conditions. Research demonstrated that the integration of SDM outputs with SSWM technologies presents significant opportunities for precision herbicide application, potentially reducing chemical inputs by 40%–60% while maintaining control efficacy (Gerhards et al. 2022). Agricultural extension services should prioritize training programs for growers in high-risk regions, emphasizing the importance of seedbank management given *D. radulans's* buried seed viability.

From a policy perspective, we recommend: 1) incorporating SDM-based risk maps into biosecurity planning and regional weed management strategies; 2) establishing mandatory reporting systems in high-risk zones to track population expansion; 3) funding research on herbicide resistance management given the risk in cotton systems (Werth et al. 2011) and 4) developing climate-adaptive weed management guidelines that account for shifting habitat suitability under future scenarios. These habitat suitability maps should be integrated into existing decision support systems to guide resource allocation and management priorities. Integration with ecological succession principles further offers additional management opportunities. In disturbed agricultural landscapes, promoting rapid establishment of competitive crop species and cover crops can exploit similar niche requirements as *D. radulans*, thereby reducing invasion success (Davis et al. 2000). The identification of soil pH (30% permutation importance) and temperature seasonality (37% importance) as key determinants enables predictive targeting of lime application in acidic soils to reduce habitat suitability while improving crop productivity—a win-win strategy for integrated management.

Conclusion

This study represents the first comprehensive effort to map and model the potential distribution of *D. radulans* across Australia. Key

environmental factors, including seasonal temperature, soil pH, and precipitation patterns, have been identified, underscoring the potential for more targeted and efficient weed control measures. *D. radulans* is an emerging weed in Australia that poses significant threats to cropping systems, particularly cotton production. This study provides the first comprehensive, multimodel assessment of *D. radulans* habitat suitability across Australia, integrating field surveys with historical records to develop robust predictive models. Our findings demonstrate that seasonal temperature (bio4) and soil pH (soil1) are critical determinants of distribution, collectively contributing up to 67% of permutation importance across models. Among the four machine learning approaches evaluated, Maxent exhibited superior predictive accuracy (AUC = 0.90) and greater adaptability to varying threshold requirements, though ensemble approaches may further enhance prediction reliability. The high adaptability of *D. radulans* to diverse climatic conditions presents a significant risk for establishment in new regions, a concern previously noted by Asaduzzaman et al. (2018). The vulnerability of specific areas, such as southeastern Australia, where the species may thrive, thus supporting efforts to increase awareness and provide targeted education and training for local farmers. As cotton is a primary crop within the surveyed region, fostering sustainable agricultural practices is crucial. Raising awareness among growers, policymakers, and the public is essential to effectively manage this resilient small-seeded weed.

Limitations and future directions: By integrating ground-truth data with historical records, our findings offer valuable insights for enhancing management strategies, particularly by predicting and understanding the *D. radulans'* spread across varied environmental contexts. However, several constraints should be acknowledged these includes the sampling bias remains inherent in GBIF datasets, with potential underrepresentation in remote regions despite our spatial thinning protocols (Beck et al. 2014). The reliance on AUC and TSS is appropriate; however, these metrics alone may be insufficient. Future analyses should consider incorporating additional evaluation measures (e.g., Kappa, specificity) to provide a more comprehensive assessment. The biotic interactions, such as competition, herbivory, and soil microbial communities, were not explicitly modelled, which may influence realized distributions beyond our abiotic predictions (Wisz et al. 2013). The temporal dynamics in land use and agricultural intensification were not incorporated, though these significantly affect weed establishment opportunities (Booth and Swanton 2002). Future research on *D. radulans* should incorporate climate change projections under Representative Concentration Pathways (RCPs) and Shared Socioeconomic Pathways scenarios to assess shifting habitat suitability in southeastern Australia and validate model predictions with independent field surveys. Integrating crop rotation and management intensity as predictors, alongside temporal analyses of population spread, would enhance model accuracy and support adaptive management. Coupled with cost–benefit assessments of model-informed strategies, these approaches can guide evidence-based policy and practical decision-making, strengthening weed management and sustainability across Australian agricultural landscapes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

CRediT authorship contribution statement

Md Asaduzzaman: Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mohammad Shamim Hasan Mandal:** Writing – review & editing, Formal analysis. **Eric Koetz:** Funding acquisition, Data curation. **Azizur Rahman:** Writing – review & editing. **Sheikh Muhammad Masum:** Writing – review & editing.

Acknowledgments

We thanks to Cotton Research and Development Corporation for funding. We extend our heartfelt thanks to Professor Michael Walsh from Gulbali Institute, Charles Sturt University, for his meticulous review and constructive feedback on the initial draft of this manuscript. His insightful suggestions and expertise were instrumental in enhancing the clarity, depth, and overall quality of this article.

Appendix

A1. Area (km^2) of the suitability classes estimated based on Maxent model.

Suitability classes	Pixel count	Area (km^2)
Not suitable	9 015 687	6 549 017
Least suitable	748 743	543 888
Moderate suitable	327 023	237 550
High	180 793	131 328
Very high	170 379	123 763

References

Abdelaal, M., Fois, M., Fenu, G., Bacchetta, G., 2019. Using MaxEnt modeling to predict the potential distribution of the endemic plant Rosa arabica Crép. Ecological Informatics 50, 68–75. doi:10.1016/j.ecoinf.2019.01.003.

Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B., Anderson, R.P., 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography 38 (5), 541–545. doi:10.1111/ecog.01132.

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modeling. Journal of Biogeography 33 (10), 1677–1688. doi:10.1111/j.1365-2699.2006.01584.x.

Asaduzzaman, M., Koetz, E., 2018. Germination biology of button grass (*Dactyloctenium radulans* (R.Br.) P. Beauv.): an emerging summer grass weed in cotton farming systems. NSW DPI, Wagga Wagga. https://www.dpi.nsw.gov.au/__data/assets/pdf_file/0006/834909/SNSWRR2018-web-not-accessible-28sept2018.pdf. Accessed 15 April 2025.

Asaduzzaman, M., Koetz, E., Rahman, A., 2019. Factors affecting seed germination and emergence of button grass (*Dactyloctenium radulans* (R.Br.) P.Beauv.). Weed Biology and Management 19 (3), 85–92. doi:10.1111/wbm.12183.

Beck, J., Böller, M., Erhardt, A., Schwanghart, W., 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. Ecological Informatics 19, 10–15. doi:10.1016/j.ecoinf.2013.11.002.

Beck, J.J., Li, D., Johnson, S.E., Rogers, D., Cameron, K.M., Sytsma, K.J., Waller, D.M., 2022. Functional traits mediate individualistic species-environment distributions at broad spatial scales while fine-scale species associations remain unpredictable. American Journal of Botany 109 (12), 1991–2005. doi:10.1002/ajb2.16085.

Booth, B.D., Swanton, C.J., 2002. Assembly theory applied to weed communities. Weed Science 50 (1), 2–13. doi:10.1614/0043-1745(2002)050,[0002:AIATAT]2.0.CO;2.

Borger, C.P.D., Hashem, A., Pathan, S., 2020. Persistence of weed seedbanks and implications for long-term management in cropping systems. Weed Science 68 (1), 3–10. doi:10.1017/wsc.2019.52.

Bradley, B.A., Wilcove, D.S., Oppenheimer, M., 2010. Climate change increases risk of plant invasion in the Eastern United States. Biological Invasions 12 (7), 1855–1872. doi:10.1007/s10530-009-9597-y.

Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32. doi:10.1023/A:1010933404324.

Cameron, J., Storrie, A., 2014. Summer fallow weed management: a reference for grain growers and advisers in the southern and western grains regions of Australia. GRDC. https://www.grdc.com.au/GRDC-Manual-SummerFallowWeedManagement. Accessed 22 March 2025.

Chamberlain, S., 2021. Spocc: an R interface to species occurrence data. Journal of Open Source Software 6 (62), 3056. doi:10.21105/joss.03056.

Chauhan, B.S., Manalil, S., 2022. Seedbank persistence of four summer grass weed species in the northeast cropping region of Australia. PLoS One 17, e0262288. doi:10.1371/journal.pone.0262288.

Christensen, S., Søgaard, H.T., Kudsk, P., Nørremark, M., Lund, I., Nadimi, E.S., Jørgensen, R., 2009. Site-specific weed management. Crop Protection 28 (5), 409–417. doi:10.1016/j.cropro.2009.02.003.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88 (11), 2783–2792. doi:10.1890/07-0539.1.

Davis, M.A., Grime, J.P., Thompson, K., 2000. Fluctuating resources in plant communities: a general theory of invasibility. Journal of Ecology 88 (3), 528–534. doi:10.1046/j.1365-2745.2000.00473.x.

De'ath, G., 2007. Boosted trees for ecological modeling and prediction. Ecology 88 (1), 243–251. doi:10.1890/0012-9658(2007)88, [243:BTFEMA]2.0.CO;2.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36 (1), 27–46. doi:10.1111/j.1600-0587.2012.07348.x.

DPIRD (Department of Primary Industries and Regional Development), 2017. Weed management guidelines for Australian agriculture. https://www.agriculture.gov.au/. Accessed 7 October 2025.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29 (2), 129–151. doi:10.1111/j.2006.0906-7590.04596.x.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77 (4), 802–813. doi:10.1111/j.1365-2656.2008.01390.x.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17 (1), 43–57. doi:10.1111/j.1472-4642.2010.00725.x.

Evans, J.S., Murphy, M.A., Holden, Z.A., Cushman, S.A., 2011. Modeling species distribution and change using random forest. In: Drew, C., Wiersma, Y., Huettmann, F. (Eds.), Predictive species and habitat modeling in landscape ecology. Springer, New York, NY, pp. 139–159. doi:10.1007/978-1-4419-7390-0_8.

Everist, S.L., Poisonous plants of Australia, rev. ed. London; Sydney: Angus & Robertson.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37 (12), 4302–4315. doi:10.1002/joc.5086.

Fourcade, Y., Engler, J.O., Rödder, D., Secondi, J., 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. PLoS One 9 (5), e97122. doi:10.1371/journal.pone.0097122.

Franklin, J., 2010. Mapping species distributions: spatial inference and prediction. Cambridge University Press, London, UK.

Franklin, J., 2013. Species distribution models in conservation biogeography: developments and challenges. Diversity and Distributions 19 (10), 1217–1223. doi:10.1111/ddi.12125.

Friedl, M., Sulla-Menashe, D., 2019. MCD12Q1 MODIS/Terra+Aqua land cover type yearly L3 Global 500m SIN grid V006. NASA EOSDIS land processes distributed active archive center. https://doi.org/10.5067/MODIS/MCD12Q1.006

Friedl, M., Sulla-Menashe, D., 2022. MODIS/Terra+Aqua land cover type yearly L3 global 500m SIN grid V061. NASA EOSDIS land processes distributed active archive center. https://doi.org/10.5067/MODIS/MCD12Q1.061

Gardner, C.A., 1952. Flora of Western Australia: Volume 1, Part 1. Gramineae. Government Printer, Perth, Western Australia.

Gerhards, R., Sanchez, D.A., Hamouz, P., Peteinatos, G.G., Christensen, S., Fernandez-Quintanilla, C., 2022. Advances in site-specific weed management in agriculture—a review. Weed Research 62 (2), 123–133. doi:10.1111/wre.12526.

Ghatrehsamani, S., Jha, G., Dutta, W., Molaei, F., Nazrul, F., Fortin, M., Bansal, S., Debangshi, U., Neupane, J., 2023. Artificial intelligence tools and techniques to combat herbicide resistant weeds—a review. Sustainability 15, 1843. doi:10.3390/su15031843.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202, 18–27. doi:10.1016/j.rse.2017.06.031.

Graham, C.H., Hijmans, R.J., 2006. A comparison of methods for mapping species ranges and species richness. Global Ecology and Biogeography 15 (6), 578–587. doi:10.1111/j.1466-8238.2006.00257.x.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8 (9), 993–1009. doi:10.1111/j.1461-0248.2005.00792.x.

Hao, T., Elith, J., Guillera-Arroita, G., Lahoz-Monfort, J.J., 2019. A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. Diversity and Distributions 25 (5), 839–852. doi:10.1111/ddi.12892.

Heikkinen, R.K., Bocedi, G., Kuussaari, M., Heliölä, J., Leikola, N., Pöyry, J., Travis, J.M., 2014. Impacts of land cover data selection and trait parameterisation on dynamic modelling of species' range expansion. PLoS ONE 9 (9), e108436.

Hengl, T., Sierdsema, H., Radović, A., Dilo, A., 2009. Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. Ecological Modelling 220 (24), 3499–3511. doi:10.1016/j.ecolmodel.2009.06.038.

Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29 (5), 773–785. doi:10.1111/j.0906-7590.2006.04700.x.

Hijmans, R.J., 2020. raster: geographic data analysis and modeling. R package version 3.3-13. https://CRAN.R-project.org/package=raster. Accessed 5 November 2025.

Hirzel, A.H., Le Lay, G., 2008. Habitat suitability modelling and niche theory. Journal of Applied Ecology 45 (5), 1372–1381. doi:10.1111/j.1365-2664.2008.01524.x.

Jiménez-Valverde, A., Barve, N., Lira-Noriega, A., Maher, S.P., Nakazawa, Y., Papeş, M., Soberón, J., Sukumaran, J., Peterson, A.T., 2011. Dominant climate influences on North American bird distributions. Global Ecology and Biogeography 20 (1), 114–118. doi:10.1111/j.1466-8238.2010.00574.x.

Jiménez-Valverde, A., Lobo, J.M., Hortal, J., 2008. Not as good as they seem: the importance of concepts in species distribution modelling. Diversity and Distributions 14 (6), 885–890. doi:10.1111/j.1472-4642.2008.00496.x.

Johnson, C.J., Gillingham, M.P., 2004. Mapping uncertainty: sensitivity of wildlife habitat ratings to expert opinion. Journal of Applied Ecology 41, 1032–1041. doi:10.1111/j.0021-8901.2004.00975.x.

Johnston, A., Moran, N., Musgrove, A., Fink, D., Baillie, S.R., 2020. Estimating species distributions from spatially biased citizen science data. Ecological Modelling 422, 108927. doi:10.1016/j.ecolmodel.2019.108927.

Kass, J.M., Muscarella, R., Galante, P.J., Bohl, C.L., Pinilla-Buitrago, G.E., Boria, R.A., Soley-Guardia, M., Anderson, R.P., 2021. ENMeval 2.0: redesigned for customizable and reproducible modeling of species' niches and distributions. Methods in Ecology and Evolution 12 (9), 1602–1608. doi:10.1111/2041-210X.13628.

Lek, S., Guégan, J.F., 1999. Artificial neural networks as a tool in ecological modelling, an introduction. Ecological Modelling 120 (2–3), 65–73. doi:10.1016/S0304-3800(99)00092-7.

Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17 (2), 145–151. doi:10.1111/j.1466-8238.2007.00358.x.

Mandal, M.S.H., Okuda, T., Mon, S.M., Chew, W.C., Shigematsu, C., Hosaka, T., 2023. Habitat suitability mapping for a high-value non-timber forest product: A case study of Rauvolfia serpentina. Tropics 31 (4), 111–133. doi:10.3759/tropics.SINT04.

McKenzie, R.A., Rayner, A.C., Thompson, G.K., Pidgeon, G.F., Burren, B.R., 2004. Nitrate-nitrite toxicity in cattle and sheep grazing *Dactyloctenium radulans* (button grass) in stockyards. Australian Veterinary Journal 82 (10), 630–634. doi:10.1111/j.1751-0813.2004.tb12612.x.

Olden, J.D., Kennard, M.J., Pusey, B.J., 2008. Species invasions and the changing biogeography of Australian freshwater fishes. Global Ecology and Biogeography 17 (1), 25–37. doi:10.1111/j.1466-8238.2007.00340.x.

Peterman, W.E., Crawford, J.A., Kuhns, A.R., 2013. Using species distribution and occupancy modeling to guide survey efforts and assess species status. Journal for Nature Conservation 21 (2), 114–121. doi:10.1016/j.jnc.2012.11.005.

Peterson, A.T., Papeş, M., Soberón, J., 2011. Mechanistic and correlative models of ecological niches. European Journal of Ecology 88 (8), 1811–1823. doi:10.1890/06-2070.1.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190 (3–4), 231–259. doi:10.1016/j.ecolmodel.2005.03.026.

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. Ecography 31 (2), 161–175. doi:10.1111/j.0906-7590.2008.5203.x.

Powney, G.D., Isaac, N.J.B., 2015. Beyond maps: a review of the applications of biological records. Biological Journal of the Linnean Society 115 (3), 532–542. doi:10.1111/bij.12517.

Pulliam, H.R., 2000. On the relationship between niche and distribution. Ecology Letters 3 (4), 349–361. doi:10.1046/j.1461-0248.2000.00143.x.

Qin, A., Liu, B., Guo, Q., Bussmann, R.W., Ma, F., Jian, Z., Pei, S., 2017. Maxent modeling for predicting impacts of climate change on the potential distribution of Thuja sutchuenensis Franch., an extremely endangered conifer from southwestern China. Global Ecology and Conservation 10, 139–146. doi:10.1016/j.gecco.2017.02.004.

Radosavljevic, A., Anderson, R.P., 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. Journal of Biogeography 41 (4), 629–643. doi:10.1111/jbi.12227.

Ridgeway, G.. gbm: generalized boosted regression models. R package version 1.6-3. https://CRAN.R-project.org/package=gbm. Accessed 11 August 2025.

Rios, E.V., Sadler, J., Graham, L., Matthews, T.J., 2024. Species distribution models and island biogeography: challenges and prospects. Global Ecology and Conservation 51, e02943. doi:10.1016/j.gecco.2024.e02943.

Rossel, R.A.V., Chen, C., Grundy, M.J., Searle, R., Clifford, D., Campbell, P.H., 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. Soil Research (Collingwood, Vic.) 53 (8), 845–864. doi:10.1071/SR14366.

Sarkar, D., Jagannivsan, H., Debnath, A., Talukdar, G., 2024. A systematic review on the potential impact of future climate change on India's biodiversity using species distribution model (SDM) studies: trends, and data gaps. Ecological Modelling 489, 110601. doi:10.1016/j.ecolmodel.2023.110601.

Sharp, B.L., Peterson, A.T., Gouveia, C., 2019. Modeling species distributions: a review of recent developments and future directions. Ecological Modelling 399, 1–12. doi:10.1016/j.ecolmodel.2019.01.012.

Shaw, D.R., 2005. Remote sensing and site-specific weed management. Frontiers in Ecology and the Environment 3 (10), 526–532. doi:10.2307/3868608.

Shorewala, S., Ashfaque, A., Sidharth, R., Verma, U., 2021. Weed density and distribution estimation for precision agriculture using semi-supervised learning. IEEE Access 9, 27971–27986. doi:10.1109/ACCESS.2021.3057912.

Tingley, M.W., Koo, M.S., Moritz, C., Rush, A.C., Beissinger, S.R., 2012. The push and pull of climate change causes heterogeneous shifts in avian elevational ranges. Global Change Biology 18 (11), 3279–3290. doi:10.1111/j.1365-2486.2012.02784.x.

Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD—a platform for ensemble forecasting of species distributions. Ecography 32 (3), 369–373. doi:10.1111/j.1600-0587.2008.05742.x.

Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2019. blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods in Ecology and Evolution 10 (2), 225–232. doi:10.1111/2041-210X.13107.

Vignali, S., Barras, A.G., Arlettaz, R., Braunisch, V., 2020. SDMtune: An R package to tune and evaluate species distribution models. Ecology and Evolution 10 (20), 11488–11506.

Werth, J., Thornby, D., Walker, S., 2011. Assessing weeds at risk of evolving glyphosate resistance in Australian sub-tropical glyphosate-resistant cotton systems. Crop & Pasture Science 62 (11), 1002–1009.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Svenning, J.C., 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biological Reviews 88 (1), 15–30.

Yoccoz, N.G., Nichols, J.D., Boulinier, T., 2001. Monitoring of biological diversity in space and time. Trends in Ecology & Evolution 16 (8), 446–453. doi:10.1016/S0169-5347(01)02205-4.