LETTER TO THE EDITOR

Exploring plant characteristics for constructing a pre-border weed risk assessment for China

Fang Wang · Jing Huang · Ning Zhang · Youjun Li · Shanyong He · Junbao Wen · Liping Yin · Yibing Liang

Abstract Biological invasions has caused significant damage to the ecological environment and economy of the world. Pest risk assessment is the most cost-effective means of preventing biological invasions to identify potentially suitable indexes for constructing pre-border weed risk assessment methods for China, we screened 80 metrics derived from 53 plant characteristics known to be related to invasive alien plants in other parts of the world and tested whether these metrics differed significantly between two groups of 103 invasive alien plants and 107 non-invasive plants in China. The results showed significant differences in 30 characteristics between invasive and non-invasive plants in China. Compared to the non-invasive plant group, the invasive plant group in China had a greater proportion of (1) plants native to the Americas, (2) plants belonging to the Asteraceae family, (3) polyploid plants, and had (4) a smaller proportion of plants propagated asexually only. The 30 metrics with significant differences were selected for LASSO regression to develop a predictive model to determine how well the metrics could distinguish between invasive and non-invasive alien plants already present in China. Finally an optimal model with 18 metrics was screened out. The optimal model was able to accurately discriminate 75% of non-invasive plants and 90% of invasive plants on the test set. Therefore the present study screened a range of useful metrics for the identification of invasive plants in China, and the high discriminative power of our models indicates that the subset of 18 variables retained in the final model could be useful for establishing a pre-border invasive plant screening tool for China in the future.

F. Wang · S. He (✉)
Key Laboratory of Southwest China Wildlife Resources Conservation (Ministry of Education), China West Normal University, No. 1 Shida Road, Shunqing District, Nanchong 637002, Sichuan, China
e-mail: heshanyong2007@163.com

J. Huang · Y. Liang
Center for International Inspection and Quarantine Standards and Technical Regulations, General Administration of Customs. People's Republic of China, Beijing 100013, China

N. Zhang · Y. Li · L. Yin (✉)
Technical Center for Animal Plant and Food Inspection and Quarantine of Shanghai Customs District, 1208 Minsheng Road, Pudong New District, Shanghai 200135, China
e-mail: yinliping@hotmail.com

J. Wen (✉)
State Key Laboratory of Efficient Production of Forest Resources, Beijing Forestry University, No. 35 Qinghua East Road, Haidian District, P.O. Box 113, Beijing 100083, China
e-mail: wenjb@bjfu.edu.cn

Introduction

Biological invasion is a popular topic in ecological research and an urgent problem to be solved (González et al. 2010; Van Kleunen et al. 2010; Lowry et al. 2013). Researchers have tried to explain invasion of alien plants from various aspects, including the characteristics of successful alien plants, habitat invasibility, invasion history of the plant in other regions, and evolution of plants (Rejmánek 1996; Kolar and Lodge 2001; James and Drenovsky 2007; Van Kleunen et al. 2010). Since Baker (1974, 1965) proposed classic weed characteristics, many studies have explained and predicted plant invasions based on plant functional traits (Rejmánek and Richardson 1996; Reichard and Hamilton 1997; Goodwin et al. 1999; Thompson et al. 2001; Daehler 2003; Hamilton et al. 2005; Pyšek and Richardson 2007). However, it is challenging to find the universal characteristics of that can differentiate invasive and non-invasive alien plants. Moreover, there are very few studies on exploring the characteristics of invasive alien plants using multi-species comparison studies in China (Chen et al. 2015). Among the few studies, Chen et al. (2009) comparatively analysed 23 traits between invasive alien plants, non-invasive alien plants, and native weeds in Jiangsu, Zhejiang, and Shanghai provinces of China. Chen (2012) compared the differences in nine characteristics between invasive alien plants and naturalised alien plants in China. There are other similar multi-species comparative studies, but they were limited to small-scale geographic areas, and the comparison groups were usually a few invasive alien plants and non-invasive native plants (Chen 2016; Hu et al. 2016; Wei et al. 2017). Overall, research on the characteristics of invasive plants through multi-species comparison is still in the preliminary stage in China because the comparison groups, their geographical scale, and even the plant characteristics are not presently comprehensive enough. Status identification (i.e., invasive or non-invasive alien plants) of the alien plants is a prerequisite for multi-species comparative studies.

After years of basic research, information on the list of names, the invasive or non-invasive status, and damage evaluation of alien plants in China has become relatively more comprehensive (Xu and Qiang 2004, 2011; Wan et al. 2009; He 2012; Ma 2013; Huang 2014), which provides a basis for

conducting multi-species comparisons on a national scale. On the other hand, plant trade in China began in the early 21st century (Normile 2004). However the current existing risk assessment methods are mainly suitable for managing invasive plants that have already entered China (Yin et al. 2007). Therefore, there is an urgent need for an effective pre-border risk assessment tool to screen invasive plants before introduction.

To develop a complete and efficient invasive alien plant screening tool suitable for China, it is important to identify distinguishing characteristics that can differentiate potentially invasive and non-invasive alien plants in China. To this end, in the present study, we (1) screened a range of plant characteristics previously used to distinguish between invasive vs. non-invasive plants in other parts of the world, and (2) used a predictive model to evaluate the discriminatory power of the metrics screened.

Materials and methods

Species selection

Invasive alien plants

A total of 103 invasive alien plant species that cause loss of economic or ecological benefits at the national level were selected for this analysis from the monograph 'The Checklist of the Chinese Invasive Plants' edited by Ma (2013). This monograph ranked the risk level of invasive alien plants in China into five grades based on their occurrence and the extent and degree of damage caused by them. Among the five grades, the species that cause loss of economic or ecological benefits at the national level were classified as malignant invasion (level I) (34 species) and serious invasion (level II) (69 species). Invasive plants that cause local damage or have little or no damage were classified as local invasion (level III), general invasion (level IV), and category to be seen (level V). Since the classification is not static, especially for invasive plants not in levels I and II, it is difficult to predict whether they will cause harm at the national level or be controlled by nature in the future (Yan et al. 2014). Therefore, we selected the 103 invasive alien plants listed under levels I and II.

Non-invasive alien plants

A total of 107 non-invasive alien plant species were selected for this analysis from the monograph '*Exotic Plants in China*' (He 2012). Non-invasive alien plants were identified mainly according to the criteria that they did not cause economic or environmental harm in China (Daehler et al. 2004), and the impact of the time of introduction was also considered. The damage was mainly confirmed by referring to domestic and foreign literature and authoritative pest or invasive species databases, mainly in Crop Protection Compendium, Invasive Species Compendium, and Global Invasive Species Database.

It often takes a period of time for an invasive alien species to go through establishment to spread and outbreak after being introduced. The process is called invasion lag (Song and Xu 2004). Generally, the longer the duration since a plant was introduced into a non-native region, the higher the probability that it will become invasive (Pyšek et al. 2009a, 2009b). But on the other hand, if a plant didn't cause economic and environmental harm for a long period of time after its introduction, then it will probably not become invasive. It often takes decades for an alien plant to become invasive (Williamson 1996), but the exact number of years is unknown. Relevant research in Japan has used alien plants that have been introduced for at least 40 years (Nishida et al. 2009), and at least 75 years in the United States (Koop et al. 2012). This study selected alien plants for at least 75 years from being introduced to China to ensure accuracy.

Characteristic selection

The indices set in the existing weed risk assessment methods were derived from the discriminative characteristics of invasive plants (see the indices in the Appendix 1). At present, the Australian Weed Risk Assessment method (AWRA) is widely used globally (Gordon et al. 2008; Dawson et al. 2009; Nishida et al. 2009; Crosti et al. 2010; Gassó et al. 2010; McClay et al. 2010). The AWRA has a set of 49 evaluation indexes, including the historical, biological and ecological characteristics of invasive plants, which are very comprehensive and applicable (Pheloung et al. 1999). It has also significantly influenced weed risk assessment methods in other regions, such as the United States (Koop et al. 2012). Meanwhile

the Central European weed risk assessment methods use the characteristics of invasive plants that are easy to query and evaluate as evaluation indices (Weber and Gut 2004).

To quickly obtain the characteristics needed to establish a risk assessment method for invasive plants in China, we sorted out 45 discriminative characteristics related to invasive plants included in the AWRA and the Central European weed risk assessment (see Appendix 1). Among them, 43 characteristics were selected from the AWRA, two were included in the Central European weed risk assessment, and one was included in both (Pheloung et al. 1999; Weber and Gut 2004). In order to obtain more and easily assessable invasive plant evaluation indicators, 8 characteristics that were easily obtained and related to invasive plants, were obtained from other related studies (Reichard and Hamilton 1997; Pyšek and Richardson 2007; Qiang 2009; Chen 2012). Overall, we selected a total of 53 plant characteristics, among them, 52 were qualitative characteristics, and one was quantitative. The qualitative characteristics were mainly measured using the binary method of yes/no. The details of each characteristic are included in Appendix 1. According to the nature of the characteristics, each characteristic was classified into seven sections: colonisation (9 characteristics), reproduction (12 characteristics), dispersal (8 characteristics), undesirable characteristics (10 characteristics), life form and biological habits (6 characteristics), morphology and taxonomy (3 characteristics), and weed history (5 characteristics).

Information source

The information on each characteristic for each alien plant was mainly compiled from scientific publications. When the information was contradictory or ambiguous, it was judged by the consulting experts. Alien plant species with no information on each of the characteristics was not evaluated. All searches were completed in July 2021. Cnki (www.cnki.net/), Springer (link.springer.com/), Wiley Online Library (onlinelibrary.wiley.com), and Google Scholar (scholar.google.com/) were mainly used to download the literature and books on corresponding plants. Online databases or Google search were used to directly query the relevant characteristics of plants. The Search keywords follow the

tips given in the reference of Gordon et al. (2010). For example, to find out whether a plant was shade tolerant, we used the plant name + 'shade' / 'sun' / 'light' as keywords to search online. To find out whether a plant was capable of natural hybridization, we used the plant name + 'hybrid'/'crossing' as keywords. Online plant information websites mainly included the following: Crop Protection Compendium (www.cabi.org/cpc), Invasive Species Compendium (www.cabi.org/isc), China National Pest and Quarantine Information System (www.pestchina.com), Global Invasive Species Database (www.iucngisd.org), Hawaii-Pacific Weed Risk Assessment(www.hear.org/plants/), Plant For A Future (www.pfaf.org), Seed Information Database (data.kew.org/sid/), Index to Plant Chromosome Numbers (www.tropicos.org), Useful Tropical Plants Database (tropical.theferns.info), US Department of Agriculture (plants.usda.gov), US Food and Drug Administration Toxic Plants Database (www.accessdata.fda.gov), Canadian Poisonous Plants Information System (www.cbif.gc.ca), Global Compendium of Weeds (www.hear.org/gcw/), and China Online Flora (www.iplant.cn).

Characteristic comparison and metrics selection

In order to identify metrics that can be used to formulate a predictive model for alien invasive plants, some characteristics were subdivided into multiple binary indicators, such as the area of origin, which was subdivided into five metrics (America, Europe, Asia, Africa, Oceania; Table 1 and Appendix 1). This approach yielded 80 metrics from the 53 characteristics. The 80 metrics were compared individually to explore whether there were differences between the two groups of alien plants (103 invasive alien plants and 107 non-invasive alien plants) in China. For qualitative characteristics, the chi-square test was used to test for differences (Lake and Leishman 2004; Sutherland 2004), while Fisher's exact test was used when the conditions did not meet the requirements in the Chi-square test. For quantitative characteristics, an independent sample t-test was used to test for differences.

All the above analyses were performed using the statistical analysis software SPSS Statistics V22.0 (Li and Zhang 2015). To reduce type-I error, we performed Benjamini and Hochberg correction (BH) and

Table 1 Comparison results of colonisation characteristics between invasive alien plants and non-invasive alien plants in China

NO.	Characteristic/metrics		χ^{2*}	P	The number and percentage of species that meet the characteristics**	Sample amount
1	Area of origin	America	44.176	< 0.0001	78(76%):32(30%)	103:107
		Europe	12.139	< 0.0001	7(7%):26(24%)	103:107
		Asia	14.550	0.00014	8(8%):30(28%)	103:107
		Africa	2.298	0.157	10(10%):18(17%)	103:107
		Oceania		1	0(0%):1(1%)	103:107
2	Domesticated		29.436	< 0.0001	4(6%):27(50%)	65:54
3	Inhabiting risky habitats		22.471	< 0.0001	66(66%):28(31%)	100:89
4	Shade tolerance		6.380	0.013	29(35%):14(18%)	83:80
5	Naturalised beyond native range			0.055	103(100%):94(96%)	103:98
6	Herbicides resistance			0.189	7(11%):0(0%)	64:20
7	Tolerant to low nutrients			1	32(94%):27(93%)	34:29
8	Tolerating or benefiting from disturbance		0.001	1	43(80%):27(79%)	54:34
9	Repeated introductions Outside its natural range		–	–	103(100%):107(100%)	103:107

*Null under this column means that the characteristic/metrics was tested by Fisher's exact test. Those marked as '–' under this column' means the percentage of species meeting the requirements in both invasive alien plants and non-invasive alien plants was all 0% or 100%. The heading is the same from Tables 1, 2, 3, 4, 5, 6 and 7.

**In the numbers below the column, the colon is preceded by invasive plants and followed by non-invasive plants. The heading is the same from Tables 1, 2, 3, 4, 5, 6 and 7.

Holm correction to correct the *P*-values after Chi-square comparisons (Tredennick et al. 2021). So only metrics with both *P* values < 0.05 were considered to show significant differences between invasive vs. non-invasive plants in China, and these metrics were then used in the predictive modelling below.

Formulating a predictive model for alien invasive plants in China

The metrics analyzed in this study were all discrete and binary data; therefore, Lasso regression analysis was used to screen variables and build a predictive model for alien invasive plants in China (Tredennick et al. 2021). Lasso regression analysis was performed using STATA 16. LASSO (Least absolute shrinkage and selection operator) is a regularization method that adds a penalty term to the model estimation to compress too small regression coefficients to zero at the cost of estimation bias, thus obtaining higher model prediction accuracy and model generalization capability. LASSO adds an l_1-parametric penalty term to ordinary least squares (OLS), which allows some regression coefficients to be compressed to zero and thus eliminates some variables from the model. Thus, LASSO regression can perform both model fitting and variable selection. In addition, LASSO also avoids the problems of overfitting and multicollinearity associated with OLS estimation when there are too many predictor variables (Tibshirani 1996; Tredennick et al. 2021; Zhang et al. 2020). Model fitting and variable selection in STATA is usually carried out by calculating the value of the parameter λ of the penalty function for model determination. There are three specific calculation methods including cross-validation (CV), adaptive lasso and plugin lasso. The most used method at present is cross-validation. The criterion of cross-validation is to minimize the CV function $f(λ)$, an estimate of the out-of-sample prediction error. The model for the λ that minimizes the CV function is the selected model. In this study, the CV method is also used for model selection, choosing the λ that minimize the CV function to determine the final model. LASSO is able to calculate the goodness of fit of every model simulated. The larger the CV deviation ratio and the smaller deviation indicate a better fit (StataCorp. 2021, Obuchi and Kabashima 2016).

A total of 210 species of alien plants (including 103 invasive and 107 non-invasive species) were collected in this study, and 70% (72 invasive and 75 non-invasive species) were used as the training set, while 30% served as the test set. When setting parameters in STATA 16, the Logit model was selected, the random seed number was set to 123, and the 10-fold cross-validation method was employed to construct multiple potential predictive models. Predictive models should be able to calculate the probability of a plant species becoming invasive. We classified species with a probability > 0.5 as invasive, and a probability ≤ 0.5 as non-invasive (Tang and Li 2014). We determined the model selected by the software that minimize the CV function to be the final optimal model and tested the evaluation outcome of the optimal model on the test set. Additionally, a receiver operating characteristic (ROC) curve was used to test the evaluation outcome of the prediction model.

Results

Colonisation

Among the nine characteristics classified as colonisation, in terms of repeated introduction outside their natural range, all 210 selected plants had a history of repeated introduction. There was no significant difference in this characteristic between invasive alien plants and non-invasive alien plants in China (Table 1). Among the remaining eight characteristics, there were significant differences in the characteristics of the source area, domestication, inhabiting risky habitats, and shade tolerance between invasive alien plants and non-invasive alien plants in China (Table 1). Among them, in the invasive alien plant group, the proportion of plants that were native to the Americas, not domesticated, inhabiting the risky habitats, and shade tolerance was significantly higher than that of the non-invasive alien plant group (Table 1). In the non-invasive alien plant group, the proportion of plant species native to Europe and Asia was significantly higher than that in the invasive alien plant group (Table 1).

Reproduction

Among the 12 characteristics classified into reproduction, there were no significant differences between the invasive alien plant and the non-invasive alien plant groups in the three characteristics of seed germination requirements, specialist pollinators and substantial reproductive failure in native habitats (Table 2). None of the 92 selected plants (49 invasive alien plants and 43 non-invasive alien plants) failed to breed for the latter characteristic. There were significant differences between the two groups of plants for the remaining ten characteristics. In the invasive alien plant group, the proportion of plant species with a persistent propagule bank, polyploidy, high reproductive capacity, self-fertilisation, and natural hybridisation was significantly greater than that in the non-invasive alien plant group (Table 2). In terms of the characteristics of minimum generative time, the proportion of plants with a generative time less than one year in the invasive alien plant group was significantly greater than that in the non-invasive alien

plant group (Table 2). In terms of the characteristic of pollinators, the proportion of plants with wind-pollinated plants in the invasive alien plant group was significantly higher than that in the non-invasive plant group (Table 2). There were no significant differences between the two groups in other pollination methods. In terms of reproductive mode, the proportion of plants with only asexual reproduction was significantly less in the invasive alien plant group than that in the non-invasive alien plant group (Table 2). The proportion of plants with both asexual and sexual reproduction in the invasive alien plant group was significantly higher than that in the non-invasive alien plant group (Table 2). The two groups of plants had no significant differences in the characteristic of only sexual reproduction. With regard to the plant reproductive system, the proportion of polygamo-monoecious plants in invasive alien plants was significantly higher than in non-invasive alien plants (Table 2). However, there was no significant difference between the two groups in terms of the characteristics of hermaphrodite, dioecious, and polygamo-dioecious.

Table 2 Comparison results of reproduction characteristics between invasive alien plants and alien non-invasive plants in China

No.	Characteristic/metrics		χ^2	P	The number and percentage of species that meet the characteristics	Sample amount
1	With persistent propagule bank		42.894	<0.0001	78(94%):33(46%)	83:71
2	Minimum generative time (one year)		32.808	<0.0001	94(94%):54(59%)	100:91
3	Polyploidy		22.604	<0.0001	47(55%):16(19%)	86:83
4	Pollinators	Wind-pollinator	12.230	0.001	28(35%):13(13%)	80:100
		Insect-pollinator	2.880	0.112	48(60%):72(72%)	80:100
		Multiple pollinators	1.549	0.269	4(5%):10(10%)	80:100
		Bird-pollinator		0.067	0(0%):5(5%)	80:100
5	Specialist pollinators			0.060	0(0%):5(6%)	79:87
6	Reproductive mode	Only vegetative propagation	15.155	<0.0001	1(1%):16(16%)	101:97
		Both	9.477	0.003	40(40%):19(20%)	101:97
		Only sexual reproduction	0.426	0.560	60(59%):62(64%)	101:97
7	Prolific production		13.327	0.0004	29(59%):8(21%)	49:39
8	Self-fertilisation		10.960	0.001	65(84%):46(66%)	77:76
9	Hybridising naturally		5.069	0.032	26(81%):17(55%)	32:31
10	Reproductive system	Polygamo-monoecious	7.785	0.006	12(17%):3(4%)	72:85
		Monoecious	3.55	0.079	60(83%):79(93%)	72:85
		Dioecious		0.5	0(0%):2(2%)	72:85
		Polygamo-dioecious		1	0(0%):1(1%)	72:85
11	Seed germination requirements		0.122	0.857	36(47%):23(44%)	76:52
12	Substantial reproductive failure in native		–	–	0(0%):0(0%)	49:43

Dispersal

There were significant differences between the invasive alien plant group and the non-invasive alien plant group in China in the eight dispersal characteristics (Table 3). Among them, in terms of the characteristic of 'dispersed intentionally by people', the proportion of plants that can be actively spread by people in the invasive alien plant group was significantly lower than that in the non-invasive alien plant group (Table 3). For each of the remaining seven characteristics, the proportion of plants in the invasive alien plant group was significantly higher than that in the non-invasive alien plant group (Table 3).

Undesirable characteristics

Among the ten undesirable characteristics, the invasive alien plant group and non-invasive alien plant group in China did not show significant differences in the four characteristics causing fire hazards, parasitic plants, unpalatable to grazing animals climbing or covering growth habit (Table 4). The proportion of plants that can carry important pests and diseases in the invasive alien plant group in China was significantly lower than that in the non-invasive alien plant group (Table 4). For each of the remaining five characteristics that were of forming dense thickets, having allelopathic effects, being toxic to animals, being allergic or toxic to humans, and having thorns, spines or burrs, the proportion of plants in the invasive alien plant group was significantly higher than that in the non-invasive alien plant group (Table 4).

Life form and biological habits

There was no significant difference between the proportion of plants in the invasive alien plant group and

Table 3 Comparison results of dispersal characteristics between invasive alien plants and alien non-invasive plants in China

No.	Characteristic/metrics	χ^2	P	The number and percentage of species that meet the characteristics	Sample amount
1	Dispersed as contaminants of produce	87.882	< 0.0001	65(81%):8(9%)	80:89
2	Dispersed unintentionally by people	77.164	< 0.0001	82(89%):25(26%)	92:97
3	Water dispersal	74.776	< 0.0001	78(90%):15(21%)	87:70
4	Dispersed by other animals externally	69.668	< 0.0001	58(73%):11(11%)	809:98
5	Wind dispersal	50.384	< 0.0001	56(73%):16(18%)	77:89
6	Dispersed intentionally by people	29.360	< 0.0001	75(76%):107(100%)	99:107
7	Bird dispersal	14.789	< 0.0001	34(65%):22(31%)	52:72
8	Dispersed by other animals internally	12.040	0.001	33(79%):27(44%)	42:61

Table 4 Comparison results of undesirable characteristics between invasive alien plants and alien non-invasive plants in China

No.	Characteristic/metrics	χ^2	P	The number and percentage of species that meet the characteristics	Sample amount
1	Forming dense thickets	63.262	< 0.0001	63(76%):12(14%)	83:83
2	Allelopathic	46.7	< 0.0001	38(68%):7(10%)	56:72
3	Toxic to animals	24.480	< 0.0001	44(52%):14(16%)	85:87
4	Allergic or toxic to humans	17.028	< 0.0001	37(50%):17(19%)	74:88
5	Host for recognised pests and pathogens	14.782	0.0003	55(53%):84(79%)	103:107
6	Producing spines, thorns, or burrs	6.992	0.011	22(21%):9(8%)	103:107
7	Creating a fire hazard in natural ecosystems		0.068	6(15%):2(4%)	40:54
8	Parasitic		0.488	1(1%):0(0%)	103:107
9	Unpalatable to grazing animals	0.976	0.344	26(36%):15(28%)	72:54
10	Climbing or smothering growth habit	0.013	1	15(15%):15(14%)	103:107

the non-invasive alien plant group in China in terms of nitrogen fixation characteristics, aquatic and geophyte (Table 5). In each of the three characteristics of life form, flowering characteristics, and fruiting characteristics, the proportion of plants in the group of invasive alien plants was significantly different from that of non-invasive alien plants. In terms of life form, the proportion of small herbaceous perennials in the invasive alien plant group was significantly lower than that in the non-invasive alien plant group (Table 5). The proportion of small herbaceous plants was significantly higher than that in the non-invasive alien plant group (Table 5). In terms of flowering characteristics, the proportion of plants beginning flowering in spring in the invasive alien plant group was significantly lower than that in the non-invasive alien plant group (Table 5). In addition, the proportion of plants that bloomed throughout the year in the invasive alien plant group was significantly higher than that in the non-invasive plant group. In terms of fruiting characteristics, the proportion of plants that bear fruit throughout the year among the invasive alien plants was significantly higher than that in the non-invasive alien plants (Table 5).

Morphology and taxonomy

It showed that there were significant differences between the invasive alien plant group and the non-invasive alien plant group in China in each of the three characteristics of fruit or seed morphology, seed quality, and taxonomy (Table 6). In terms of the fruit morphology, the proportion of plants whose fruit or seed had long-distance dispersal structures in the invasive alien plant group were significantly higher than that in the non-invasive alien plant group (Table 6). Additionally, the proportion of plants with large fleshy fruits was significantly lower than that in the non-invasive alien plant group. In terms of seed quality, the average 1000-seed mass of invasive alien plants (12.66 g) was significantly lower than that of non-invasive alien plants (242.10 g) (Table 6). In terms of classification, the number of plants belonging to Asteraceae and Poaceae in invasive alien plants was significantly higher than that in the non-invasive alien plants (Table 6).

Table 5 Comparison results of life history and life form between invasive alien plants and alien non-invasive plants in China

No.	Characteristic/metrics		χ^2	P	The number and percentage of species that meet the characteristics	Sample amount
1	Life form	Woody perennial	5.595	0.020	12(11%):25(22%)	114:113
		Large annual	4.762	0.038	39(34%):24(21%)	114:113
		Small annual	2.929	0.116	25(22%):15(13%)	114:113
		Small herbaceous perennial	2.711	0.109	14(12%):23(20%)	114:113*
		Large herbaceous perennial	0.126	0.751	24(21%):26(23%)	114:113
2	Fixing nitrogen		1.815	0.211	11(11%):6(6%)	103:107
3	Aquatic			0.062	6(6%):1(1%)	103:107
4	Geophyte		2.910	0.101	13(13%):23(21%)	103:107
5	Flowering season	Spring	6.342	0.013	29(30%):47(47%)	98:100
		Summer	1.623	0.248	45(46%):37(37%)	98:100
		Fall	0.313	0.767	5(5%):7(7%)	98:100
		Winter		1	2(2%):3(3%)	98:100
		All year	6.207	0.014	17(17%):6(6%)	98:100
6	Fruiting season	Spring	0.065	0.844	16(21%):16(22%)	78:72
		Summer	0.50	0.870	43(55%):41(57%)	78:72
		Fall	1.409	0.316	7(9%):11(15%)	78:72
		Winter		0.229	0(0%):2(3%)	78:72
		All year	7.032	0.010	12(15%):2(3%)	78:72

*Repeat counting because some plants could be both perennial and annual

Table 6 Comparison results of morphology and taxonomy between invasive alien plants and alien non-invasive plants in China

No.	Characteristic/metrics		χ^2	P	The number and percentage of species that meet the characteristics	Sample amount
1	Fruit or seed morphology	Fruit or seed having well-developed structures for long-distance dispersal	25.015	<0.0001	58(56%):22(22%)	103:100
		Fleshy fruit and larger than 5 cm in diameter		0.007	0(0%):8(8%)	97:100
		Fleshy fruit and smaller than 5 cm in diameter	0.029	1	9(9%):10(10%)	97:100
2	Thousand-seed weight *			0.0001		93:82
3	Taxonomy	Asteraceae	18.007	<0.0001	28(27%):6(6%)	103:107
		Poaceae	5.566	0.022	13(13%):4(4%)	103:107

*An independent sample t-test was used

Weed history

In each of the four characteristics, weeds of agriculture/ horticulture/ forestry, environmental weed, garden/ amenity/ disturbance weed, and congeneric weed, the proportion of plants possessing any one of the four characteristics in the group of invasive alien plants in China was significantly higher than that in the group of non-invasive alien plants (Table 7). The proportion of plants with weedy races in the group of invasive alien plants in China was not significantly different from that in the non-invasive alien plant group (Table 7).

Metrics selection

Two metrics without significant differences from Colonisation and Reproduction were excluded from metrics selection (i.e., 'repeated introductions outside its natural range' and 'substantial reproductive failure in native'). Since almost all of the metrics were binary,

we also excluded thousand-seed weight from the metrics selection. So the remaining 77 metrics were performed Benjamini and Hochberg correction (BH) and Holm correction (Holm). Finally 30 metrics of the 77 metrics were identified as significant differences by both BH and Holm correction between the invasive and non-invasive plant groups (The top 30 metrics in Table 8), with five of the metrics from colonisation, eight from dispersal, two from morphology and taxonomy, six from reproduction, five from Undesirable characteristics and four from weed history (Table 8).

Optimal predictive model selection

LASSO simulated 100 models with 10-fold cross-validation to assess model accuracy. The optimal prediction model for alien invasive plants in China was identified as having $\lambda = 0.018$ and 18 coefficients. This model parameterisation minimised the CV function (Fig. 1), and yielded the minimum CV deviation and maximum dev. Ratio (Table 9). The optimal

Table 7 Comparison results of associated characteristics between invasive alien plants and alien non-invasive plants in China

No.	Characteristic/metrics	χ^2	P	The number and percentage of species that meet the characteristics	Sample amount
1	Weed of agriculture/ horticulture/ forestry	122.110	<0.0001	83(87%):5(6%)	95:88
2	Environmental weed	47.582	<0.0001	49(64%):10(12%)	77:86
3	Garden/ amenity/ disturbance weed	53.017	<0.0001	50(63%):7(8%)	80:84
4	Congeneric weed	27.853	<0.0001	78(76%):42(40%)	103:106
5	Having weedy races	3.759	0.069	3(13%):12(36%)	23:33

Table 8 Metrics selected by Benjamini and Hochberg correction (BH) and Holm correction

No.	Characteristic	Metrics*	P	p.adjust (BH)	p.adjust (Holm)
1	Weed history	Weed of agriculture/horticulture/forestry	4.664E-32	3.591E-30	3.591E-30
2	Dispersal	Dispersed as contaminants of produce	1.479E-22	5.695E-21	1.124E-20
3	Dispersal	Dispersed unintentionally by people	1.659E-19	4.258E-18	1.244E-17
4	Dispersal	Water dispersal	4.908E-19	9.447E-18	3.632E-17
5	Dispersal	Dispersed by other animals externally	1.346E-17	2.072E-16	9.824E-16
6	Undesirable characteristics	Forming dense thickets	6.224E-16	7.988E-15	4.481E-14
7	Weed history	Garden/amenity/disturbance weed	7.002E-14	7.702E-13	4.971E-12
8	Dispersal	Wind dispersal	6.080E-13	5.852E-12	4.256E-11
9	Weed history	Environmental weed	2.244E-12	1.920E-11	1.548E-10
10	Undesirable characteristics	Allelopathic	4.075E-12	3.138E-11	2.771E-10
11	Colonisation(origin)	America	1.991E-11	1.394E-10	1.334E-09
12	Reproduction	With persistent propagule bank	2.000E-11	1.283E-10	1.320E-09
13	Dispersal	Dispersed intentionally by people	4.446E-09	2.633E-08	2.890E-07
14	Reproduction	Minimum generative time	4.792E-09	2.636E-08	3.067E-07
15	Colonisation	Domestic	4.295E-08	0.000000	0.000003
16	Weed history	Congeneric weed	1.656E-07	0.000001	0.000010
17	Morphology and Taxonomy	Fruit or seed having well-developed structures for long-distance dispersal	0.000001	0.000004	0.000048
18	Undesirable characteristics	Toxic to animals	0.000001	0.000004	0.000053
19	Colonisation	Inhabiting risky habitats	0.000003	0.000012	0.000177
20	Reproduction	Polyploidy	0.000003	0.000012	0.000177
21	Morphology and Taxonomy	Asteraceae	0.000022	0.000081	0.001254
22	Undesirable characteristics	Allergic or toxic to humans	0.000051	0.000179	0.002856
23	Reproduction	Vegetative propagation	0.000060	0.000201	0.003300
24	Colonisation	Asia	0.000135	0.000433	0.007290
25	Dispersal	Bird dispersal	0.000136	0.000433	0.007290
26	Undesirable characteristics	Host for recognised pests and pathogens	0.000141	0.000433	0.007332
27	Reproduction	Prolific production	0.000430	0.001226	0.021930
28	Colonisation	Europe	0.000533	0.001466	0.026650
29	Dispersal	Dispersed by other animals internally	0.000568	0.001508	0.027832
30	Reproduction pollinator	Wind-pollinator	0.000616	0.001581	0.029568
31	Reproduction	Self-fertilisation	0.001084	0.002693	0.050948
32	Reproduction	Both	0.002984	0.007180	0.137264
33	Reproduction reproductive mode	Polygamo-monoecious	0.006276	0.014644	0.282420
34	Morphology and Taxonomy	Fleshy fruit and larger than 5 cm in diameter	0.006786	0.015368	0.298584
35	Life form and Biological habits	Fruiting all year	0.010043	0.022095	0.431849
36	Undesirable characteristics	Producing spines, thorns, or burrs	0.010814	0.023130	0.454188
37	Colonisation	Shade tolerance	0.013000	0.027054	0.533000
38	Life form and Biological habits	Spring	0.013304	0.027054	0.533000
39	Life form and Biological habits	Flowering all year	0.014748	0.029118	0.575172
40	Life form and Biological habits	Woody perennial	0.020014	0.038527	0.760532
41	Morphology and Taxonomy	Poaceae	0.022431	0.042127	0.829947
42	Reproduction	Hybridising naturally	0.031943	0.058562	1
43	Life form and Biological habits	Large annual	0.037669	0.067454	1

*Only the metrics with $P < 0.05$ were listed in this table

Table 9 Predictive models simulated by Lasso

Model No.	λ	No. of nonzero coef	CV mean deviance	Out-of-sample dev. ratio
1 ~ 28	0.345625 ~ 0.0280347	0 ~ 17	1.390633 ~ 0.6133024	-0.0034 ~ 0.5575
29	0.0255442	18	0.6099088	0.5599
30	0.0232749	18	0.607274	0.5618
31	0.0212072	18	0.604779	0.5636
32	0.0193232	18	0.6024046	0.5653
33	0.0176066	18	0.6011761	0.5662
34	0.0160425	19	0.601245	0.5662
35	0.0146173	19	0.6039171	0.5642
36	0.0133187	19	0.6086411	0.5608
37	0.0121355	19	0.6154839	0.5559
38 ~ 100	0.0110575 ~ 0.0000346	20 ~ 30	0.6218679 ~ 9.147925	0.5513 ~ -5.6008

λ: It's the tuning parameter used to control the degree of compression of the regression coefficients, and the larger the value, the stronger the penalty

coef.: The regression coefficients

CV: Cross-validation

Dev: Deviance, a measure-of-fit statistic for linear and nonlinear likelihood-based models

Table 10 Predictor variables and the coefficient estimated of the optimal predictive model selected

No.	Predictor variable	Coefficient estimated*
1	Domesticated	− 0.746
2	Forming dense thickets	0.717
3	Minimum generative time (one year)	0.548
4	Polyploidy	0.510
5	With persistent propagule bank	0.474
6	Toxic to animals	0.421
7	Dispersed unintentionally by people	0.376
8	Inhabiting risky habitats	0.343
9	Water dispersal	0.315
10	Allergic or toxic to humans	0.300
11	Dispersed as contaminants of produce	0.295
12	Dispersed by other animals externally	0.236
13	Host for recognised pests and pathogens	-0.194
14	Only vegetative propagation	-0.132
15	Dispersed by other animals internally	0.117
16	Wind-pollinator	0.109
17	Area of origin (America)	0.094
18	Allelopathic	0.045

*Positive or negative coefficients represented positive or negative correlations with invasive plants, and the magnitude of the coefficient represented the degree of correlation

model could accurately identify 97% of non-invasive plants and 94% of invasive plants in the training set. The 18 predictor variables included in the optimal LASSO model and their coefficients are listed in Table 10. A positive or negative coefficient indicated

the metric had a corresponding positive or negative correlation with a plant being scored as invasive. The magnitude of the coefficient represented the degree of correlation. The metric 'domesticated' meant it had the highest negative correlation with invasive plant, and the metric 'allelopathic' had the lowest positive correlation with invasive plant.

Prediction outcome of the optimal predictive model

Using the test set, 90% of invasive plants and 75% of non-invasive plants were accurately identified by the optimal predictive model (Table 11). ROC results showed that the prediction outcomes of the model using the test set were significantly higher than 0.5 ($P < 0.05$), and the area under the curve (AUC) values were 0.925 (Fig. 2).

Discussion

Characteristics with no significant differences

We found no significant differences in 36 metrics between invasive alien plants and non-invasive alien plants in China through comparative analysis before Benjamini and Hochberg correction (BH) and Holm correction. These 36 metrics come primarily from colonisation, reproduction, undesirable characteristics, life form and biological habits.

In summary, the reasons for this result may lie in the aspects of uses of plants, life form,

Fig. 1 Different values of lambda and their corresponding values of CV function. λ: It's the tuning parameter used to control the degree of compression of the regression coefficients, and the larger the value, the stronger the penalty. λ cv : The lambda value when minimizing the value of CV function
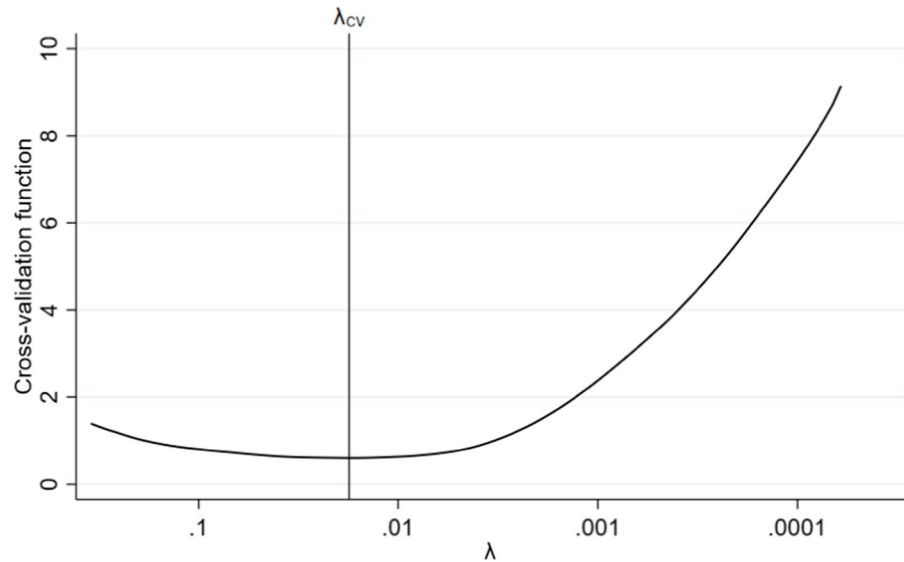

Table 11 Predictive results on test sets by the optimal predictive model

Predicted category	Priori category*	
	IPs	NIPs
IPs	90% (28/31)	25% (8/32)
NIPs	10% (3/31)	75% (24/32)

*NIPs means Non-invasive plants, IPs means Invasive plants
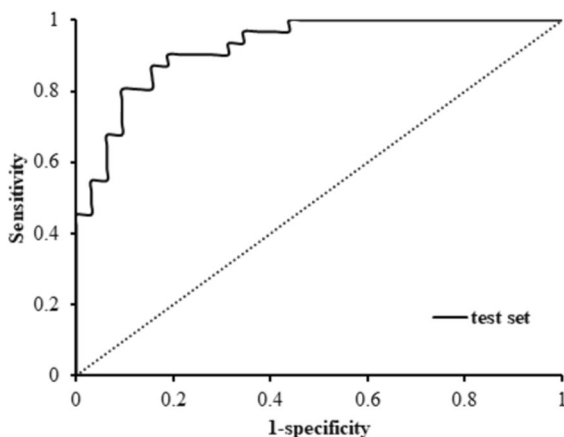
Fig. 2 ROC curve for predictive outcome of optimal predictive model selected on test set

habitat, number of samples, and information quality of alien plants in China. Concretely, there were five

characteristics in the category of colonization that showed no significant differences between the two groups of plants: 'naturalised beyond native range', 'repeated introductions outside natural range', 'herbicide resistance', 'tolerates or benefits from mutilation, cultivation, or fire', and 'low-nutrient tolerance'. The reason for the two characteristics 'naturalised beyond native range' and 'repeated introductions outside its natural range' might be related to the fact that the alien plants were used for agricultural planting, ornamental, and medicinal purposes (Weber et al. 2008). Such plants were widely circulated globally and could be introduced repeatedly in one place. At the same time, continuous planting after introduction would inevitably bring high reproduction pressure to the introduced area (Gassó et al. 2010), thus promoting the escape of alien plants and causing naturalisation (Pyšek et al., 2009a, 2009b). Therefore, we inferred that these two characteristics are not accurate in discerning invasive alien plants in China.

The reasons for the other three characteristics in the category of colonization, 'herbicide resistance', 'tolerates or benefits from mutilation, cultivation, or fire', 'low-nutrient tolerance', and the characteristic 'having weedy races' in the category of weed history, the comparative analysis results were significantly affected by the lack of information. Usually, the information for invasive alien plants was not balanced with that of non-invasive plants; the former had more information than the latter (Gassó et al. 2010; McClay

et al. 2010), and the quality of the information was generally low (Kueffer et al. 2013; Pyšek et al., 2009a, 2009b). In this study, among the four characteristics mentioned above, the sample numbers of non-invasive plants were between 20 and 33. Therefore, it is necessary to strengthen related research and information accumulation for the three characteristics in the future, especially for non-invasive plants. On the other hand, it was not easy to obtain information regarding these characteristics. Thus, presently, these characteristics might not be suitable for use as an index in the weed risk assessment in China.

There were four characteristics without significant differences between the two groups of plants in the category of undesirable characteristics: 'unpalatable to grazing animals', 'creating a fire hazard in natural ecosystems', 'climbing or smothering growth habit', and 'parasitic'. The reason that there was no significant difference on 'unpalatable to grazing animals' might mainly lie in plant use. According to the specific statistical results, 32 invasive plants can be used as animal feed, and about 40 non-invasive plants usually consumed by poultry are used as food and vegetable plants. Therefore, both groups of plants have several plants that are palatable to herbivores, resulting in no significant difference in the characteristic of unpalatability. In terms of the characteristic of creating a fire hazard in natural ecosystems, there were a very small proportion of combustible plants between the two groups. This might be because invasive plants in China are mainly present in human-managed agricultural or forestry ecosystems (Weber et al. 2008), making it difficult to form fire regimes of the grass-fire cycle similar to that in the natural environment (Fusco et al. 2019). At the same time, non-invasive plants are rarely grown in the natural environment, so the two groups of plants are not different in this characteristic. There were four characteristics that showed no significant differences between the two groups of plants in life form and biological habits (i.e., the 'life form',

'aquatic', 'geophyte', 'nitrogen fixation' categories). The reason that there was no significant difference in the four characteristics in life form and biological habits, and two characteristics 'climbing or smothering growth habit' and 'parasitic' in undesirable characteristics is likely related to the life form bias of the alien plants in China. Approximately 84% of the invasive plants in China are herbaceous plants,

and among them, there are many annual herbs, few aquatic plants, and very few vines (Weber et al. 2008; Yan et al. 2014).

There were two characteristics without significant differences between the two groups of plants in the reproduction category: substantial reproductive failure in the native habitat and seed germination requirement. For the former characteristic, such a situation rarely occurs in natural conditions (Gordon et al. 2010), therefore, the two groups in this study did not show significant differences in this characteristic. For the latter characteristic, the relevant data mainly came from the Seed Information Database (SID) and many results in the database were obtained under indoor conditions, so there might be deviations from the actual situation.

New characteristics or metrics for constructing weed risk assessment

In this study, we included some new characteristics that are not assessed in the Australian or Central European WRA: 'the source area from America/Asia/Europe', 'polyploidy', 'Asteraceae', and 'wind-pollinated'. In terms of pollinators, this study found more wind-pollinated plants in the invasive alien plant group. Different studies have shown different results for wind pollination. Some have found that it was negatively related to invasion (Williamson and Fitter 1996), while others have suggested that it was positively related to invasion (Daehler 1998). Our results suggest that wind pollination is positively correlated with plant invasion. However, the pollinator data in this study, especially the pollinator information for non-invasive plants, were mainly derived from the PFAF website. Sutherland (2004) believed that pollinator information on this website is not accurate. In view of the inconsistency of pollinators in different studies and the objective reality of information acquisition, this characteristic may cause instability in predictions. Therefore, we suggest that wind pollination should be applied with caution in the WRA constructing of China.

Plant polyploidy came from the references of Qiang (2009) and Reichard and Hamilton (1997). In recent years, many studies have begun to pay attention to the correlation between polyploid and invasive plants and found that polyploid plants are more likely to be invasive and have higher invasion ability

than diploid plants (Levin 2002; Pandit et al. 2011, 2014; Te Beest et al. 2012). Our results showed that more than half (55%) of the invasive plants in China were polyploid, significantly higher than the proportion of non-invasive plants, which is consistent with the conclusions of the previous studies. Therefore, we conclude that the characteristics of plant polyploidy could be an effective predictive index.

Before performing Benjamini and Hochberg correction and Holm correction, we found that the metrics 'polygamo-monoecious' and 'annual flowering and fruiting' differed significantly between the invasive and non-invasive plants. Although the two metrics did not show significant differences between invasive and non-invasive alien plants in China after *P*-value correction and were not screened in the finally optimal predictive model in this study, we suggest that these two metrics could be used as evaluation metrics, especially for the weed risk assessment method for China in the future. There are few studies on the effect of being polygamo-monoecious for plant invasiveness. Usually, studies compare the invasiveness of hermaphrodite and dioecious plants, but these results are controversial (Pyšek and Richardson 2007). Sutherland (2004) found that the majority of invasive alien plants were hermaphrodites. Razanajatovo et al. (2016) suggested that hermaphrodite plants may facilitate plant invasion because plants presenting this characteristic could reproduce by themselves without relying on other individuals. Compared with hermaphrodites, the breeding system of polygamo-monoecious plants may be more complicated and maybe more conducive to survival and reproduction in the natural environment. Although the absolute number of polygamo-monoecious plants among the invasive alien plants in China was not high, the relative proportion was also quite high, at 15%. At present, there are very few studies testing the influence of being polygamo-monoecious in plant invasion. Therefore, the preliminary results of this study on polygamo-monoecious plants put forward an interesting hypothesis, that is, whether the characteristic of being polygamo-monoecious contributes to plant invasion. However, given the significant difference in this characteristic between invasive and non-invasive plants, this study suggests that it could be applied as an effective index in the construction of weed risk assessment.

In terms of the characteristics of the flowering and fruit periods, the results of this study showed that invasive alien plants generally have a longer flowering and fruit period than non-invasive alien plants, and there were more plants of annual flowering and fruiting in the invasive alien plant group in China. Previous studies also found that invasive alien plants had a longer flowering and fruiting period than non-invasive plants (Goodwin et al. 1999; Lake and Leishman 2004). This might be due to the long flowering and fruiting period that could increase the reproduction time and yield, thereby increasing the chance of colonisation (Baker 1974). The reason for more invasive plants flowering and fruiting annually in this study might be related to the geographical distribution pattern of invasive plants in China. Yan et al. (2014) found there were more invasive plants in the south and east, less in the north and west in China, and many invasive plants were distributed in the southwest and southeast coastal areas, especially in Fujian, Guangdong, Guangxi, Yunnan, and Taiwan. These five provinces are located in subtropical and tropical regions (Wu et al. 2006; Yan et al. 2014), so they could provide suitable environmental conditions for the plants flowering and fruiting all year round.

The prediction accuracy of the optimal prediction model and establishment of future invasive plants screening tool

Overall, the model identified in this study had high predictive accuracy on both the training set and the test set. In particular, the average prediction accuracy for invasive plants could reach more than 90% on both training and test sets, and the prediction accuracy for invasive plants only on the test set could also reach 90%, which was higher than the prediction accuracy of the AWRA and the USA weed screening tool for invasive plants in China (84% and 68%) (He et al. 2018, 2020), and also higher than the highest prediction accuracy rate of 82% in models simulated using machine learning algorithms in China by Chen et al. (2015).

The results showed that the metrics with significant differences screened in this study could be used for the construction of screening tools for invasive plants in China, and the screening tools were likely to have good predictive ability in distinguishing invasive plants. The high predictive accuracy of the optimal

model should benefit primarily from the judging metrics, as these metrics were mainly derived from the Australian weed risk assessment, and they were obtained from different studies of invasive plant characteristics around the world (Pheloung et al. 1999). At the same time, this study adds 4 new predictors that are not adopted in the AWRA and the Central European weed risk assessment, which were the origin from America, only asexual reproduction, windborne and polyploidy. These metrics reflect the distinguishing characteristics of invasive plants in China and therefore they could contribute to predict which alien plants may become invasive in the future. For example, in this study, there were a total of 110 species of alien plants from the Americas, of which 78 species had become invasive plants in China, indicating that the source areas of invasive plants in China were significantly biased. The reason behind this bias might be a combination of climatic similarities and trade. China has a similar climate type to North America (Liu and Wei 1986), and according to the continental drift hypothesis, plants from East Asia and North America should have similar genetic backgrounds, while the climate of southeastern China has similarities to that of South America (Yan et al. 2014), so alien plants whether native to North America or South America, might find a suitable place to live in China. In addition, the frequent trade between China and the USA has led to the frequent introduction of invasive alien plants into China (Callaway et al. 2006), increasing the propagule pressure (Pyšek et al. 2015) and thus giving them a better chance of surviving in China.

But meanwhile, we also found that the overall prediction accuracy of the model on the test set was lower than the prediction result on the training set. The main reason lied in the lower prediction accuracy for non-invasive plants. Only 75% of the non-invasive plants were correctly identified, while the predictive accuracy of non-invasive plants on the training set was as high as 97%. This was very similar to the assessment accuracy of non-invasive plants in China by AWRA, which was 76% (He et al. 2018). Our result is in line with the consistent assessment of AWRA, that is, assessments accuracy for non-invasive plants are generally lower than those for invasive plants (McClay et al. 2010; Koop et al. 2012). The important reason for the lower predictive accuracy for non-invasive plants by AWRA

lies in the setting of threshold division of predictive results. So by adjusting the threshold appropriately, it is possible to improve the assessment accuracy of AWRA for non-invasive plants (Nishida et al. 2009; McClay et al. 2010; Koop et al. 2012).

But for our result, there might be other two reasons for the lower predictive accuracy of the optimal model for non-invasive plants. First, non-invasive plants and invasive plants are not absolutely discrete, but are usually in a continuous state (Gordon et al. 2008). Some non-invasive plants may become invasive in the future, and, therefore, also have one or two characteristics typical of invasive plants which might have affected the final result. For example, the probability of becoming invasive for 3 of the non-invasive plants in the test set had a result value of about 0.6. But if the value of one predictor variable was canceled, (e.g., allelopathy), then the probability would be reduced to below 0.5. So such plants in the intermediate state may affect the predictive accuracy of our model for non-invasive plants. The second reason might be related to the categories classified by the predictive model. Usually a complete weed screening tool will divide the results into three categories, major invasive plant, minor invasive plant and non-invasive plant (e.g. Kato and Hata 2006; Koop et al. 2012; Křivánek and Pyšek 2006; Pheloung et al. 1999). For minor invasive plant, a re-evaluation tool can usually be used for re-screening, which would improve the predictive accuracy for non-invasive plants, because usually the re-evaluation tool only evaluates the most important prediction features (Daehler et al. 2004; Koop et al. 2012; McClay et al. 2010).

Overall, we believed the optimal prediction model had high prediction accuracy on the test set. Firstly it could accurately identify 90% of invasive alien plants and thus excellently fulfilled the primary purpose of a WRA preventing the vast majority of invasive alien plants from entering China and causing damage. Secondly the 75% predictive accuracy for non-invasive alien plants of this model was still higher than the average of 70% predictive accuracy of AWRA for non-invasive alien plants Worldwide. Finally, because of China's large land area and the resulting wide distribution range of invasive alien plant, economic losses from invasive alien plants might be greater than the economic benefits from accepting the introduction of the 25% non-invasive alien plants.

The current results show that it is feasible to use the metrics screened in this study to construct an efficient screening tool for invasive plants in China. But before formally building an invasive plant screening tool in China, we think that the following aspects should be carefully considered and resolved. (1) Considering the division of predicted results, it is recommended to divide the results into three categories (i.e., non-invasive plant, minor invasive plant and major invasive plant), and establish a secondary screening tool with the most discriminative metrics to improve the predictive accuracy for minor invasive plant to better assist management decision-making; (2) Assess the sources of uncertainty. For example, lack of information is an important type of uncertainty (Koop et al. 2012), which can be dealt with by adopting different algorithms to build relevant predictive models, such as Bayesian methods (Tredennick et al. 2021). (3) Consider using different machine learning algorithms for invasive plant screening tool building. In fact some existing studies have shown that various machine learning models have high evaluation accuracy (Elith et al. 2006; Keller et al. 2011; Chen et al. 2015). (4) Use multiple test sets to test the evaluation outcome of the final invasive plant screening tool. For example, using invasive plants from different regions or countries as test sets, such as invasive plants that have invaded the Americas or Africa.

Conclusion

Our results suggest that, among the 80 known metrics related to plant invasion, 30 can discriminate between invasive and non-invasive alien plants in China. Therefore, we could infer that not-every character known to be related to invasive plants in other regions applies to invasive alien plants in China. Invasive alien plants in China showed a particular bias in-terms of geographic origin, taxonomy, polyploidy and reproductive mode. Compared to non-invasive plants, invasive plants in China are more likely to be native to the Americas, belong to the Asteraceae family, be polyploid, and do not propagated asexually only. The predictive model simulated with these 30 metrics in the study has high predictive accuracy on both our training and test sets. These results suggest that there are specific and suitable metrics for the construction of a pre-border weed screening tool for China. Our study has laid a preliminary foundation for further research on invasive plant screening tools for China.

Data availability The datasets generated during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Appendix

See Tables 12 and 13.

Plant characteristics selected in this study

1	Colonisation	Domesticated	yes/no	1, 2
2		Naturalised beyond the native range	yes/no	1, 2
3		Shade tolerance	yes/no	1, 2
4		Inhabiting risky habitats	yes/no	3
5		Area of origin	Asia, Africa, Europe, America, Oceania, choose one of them.	4
6		Herbicides resistance	yes/no	1, 2
7		Tolerant to low nutrients	yes/no	4
8		Tolerating or benefiting from disturbance	yes/no	1, 2
9		Repeated introductions outside its natural range	yes/no	1, 2
10	Reproduction	Reproduction mode	Vegetative propagation, sexual reproduction, or both, choose one of them.	according to 1, 2
11		Prolific production	yes/no	1, 2
12		Self-fertilisation	yes/no	1, 2
13		Seed germination requirements	yes/no	5
14		With persistent propagule bank (> 1 year)	yes/no	1, 2, 4
15		Reproductive system	Monoecious, dioecious, polygamo-monoecious, polygamo-dioecious, choose one of them.	4
16		Pollinators	Four types: wind-pollinator, insect-pollinator, bird-pollinator, and multiple pollinators. Choose one of them.	According to 1, 2
17		Specialist pollinators	yes/no	1
18		Minimum generative time ≤ 1 year	yes/no	According to 1, 2
19		Hybridising naturally	yes/no	1, 2
20		Polyploidy	yes/no	5, 6
21		Substantial reproductive failure in native habitat	yes/no	1, 2
22	Dispersal	Dispersed intentionally by people	yes/no	1, 2
23		Dispersed unintentionally by people	yes/no	1, 2
24		Dispersed as contaminants of produce	yes/no	1, 2
25		Water dispersal	yes/no	1, 2
26		Wind dispersal	yes/no	1, 2
27		Bird dispersal	yes/no	1, 2
28		Dispersed by other animals externally	yes/no	1, 2
29		Dispersed by other animals internally	yes/no	1, 2

undefined(continued)

30	Undesirable characteristics	Producing spines, thorns, or burrs	yes/no	1, 2
31		Allelopathic	yes/no	1, 2
32		Parasitic	yes/no	1, 2
33		Unpalatable to grazing animals	yes/no	1, 2
34		Toxic to animals	yes/no	1, 2
35		Host for recognised pests and pathogens	yes/no	1, 2
36		Allergic or toxic to humans	yes/no	1, 2
37		Creating a fire hazard in natural ecosystems	yes/no	1, 2
38		Climbing or smothering growth habit	yes/no	1, 2
39		Forming dense thickets	yes/no	1, 2
40	Weed history	Garden/ amenity/ disturbance weed	yes/no	1, 2
41		Weed of agriculture/ horticulture/ forestry	yes/no	1, 2
42		Environmental weed	yes/no	1, 2
43		Congeneric weed	yes/no	1, 2
44		Weedy races	yes/no	1, 2
45	Life form and Biological habits	Life form	Five types: small annual (< 80 cm), large annual (> 80 cm), woody perennial, small herbaceous perennial (< 80 cm), large herbaceous perennial (> 80 cm); choose one of them.	1, 2, 3
46		fixing nitrogen	yes/no	1, 2
47		Aquatic	yes/no	1, 2
48		Geophyte	yes/no	1, 2
49		Flowering season	Spring, summer, fall, winter, all year, choose one of them.	5
50		Fruiting season	Spring, summer, fall, winter, all year, choose one of them.	4
51	Morphology and Taxonomy	Thousand seed weight	Quantitative results	7
52		Fruit or seed morphology	Three groups. The first group is whether the plant's fruit or seed has well developed structures for long-distance dispersal (like pappus, hairs, wings, spikes, thorns). The answer is yes/no. The second group is whether the fruit of the plant fleshy and smaller than 5 cm in diameter, the answer is yes/no. The third group is whether the fruit of the plant is fleshy and larger than 5 cm in diameter. The answer is yes/no.	3
53		Taxonomy	Three groups. The first group is whether the plant belongs to the Poaceae. The answer is yes/no. Three groups. The first group is whether the plant belongs to the Asteraceae. The answer is yes/no.	According to 1, 2

List of alien plants used in this study

No.	Family	Scientific name	Status
1	Amaranthaceae	*Alternanthera philoxeroides*	Nationally invasive
2	Amaranthaceae	*Alternanthera pungens*	Nationally invasive
3	Amaranthaceae	*Amaranthus albus*	Nationally invasive
4	Amaranthaceae	*Amaranthus blitum*	Nationally invasive
5	Amaranthaceae	*Amaranthus hybridus*	Nationally invasive
6	Amaranthaceae	*Amaranthus palmeri*	Nationally invasive
7	Amaranthaceae	*Amaranthus polygonoides*	Nationally invasive
8	Amaranthaceae	*Amaranthus retroflexus*	Nationally invasive
9	Amaranthaceae	*Amaranthus spinosus*	Nationally invasive
10	Amaranthaceae	*Amaranthus viridis*	Nationally invasive
11	Amaranthaceae	*Celosia argentea*	Nationally invasive
12	Amaranthaceae	*Gomphrena celosioides*	Nationally invasive
13	Araceae	*Pistia stratiotes*	Nationally invasive
14	Asteraceae	*Ageratina adenophora*	Nationally invasive
15	Asteraceae	*Ageratum conyzoides*	Nationally invasive
16	Asteraceae	*Ageratum houstonianum*	Nationally invasive
17	Asteraceae	*Ambrosia artemisiifolia*	Nationally invasive
18	Asteraceae	*Ambrosia trifida*	Nationally invasive
19	Asteraceae	*Aster subulatus*	Nationally invasive
20	Asteraceae	*Bidens alba*	Nationally invasive
21	Asteraceae	*Bidens frondosa*	Nationally invasive
22	Asteraceae	*Bidens pilosa*	Nationally invasive
23	Asteraceae	*Chromolaena odorata*	Nationally invasive
24	Asteraceae	*Crassocephalum crepidioides*	Nationally invasive
25	Asteraceae	*Erigeron annuus*	Nationally invasive
26	Asteraceae	*Erigeron bonariensis*	Nationally invasive
27	Asteraceae	*Erigeron canadensis*	Nationally invasive
28	Asteraceae	*Erigeron sumatrensis*	Nationally invasive
29	Asteraceae	*Flaveria bidentis*	Nationally invasive
30	Asteraceae	*Galinsoga parviflora*	Nationally invasive
31	Asteraceae	*Galinsoga quadriradiata*	Nationally invasive
32	Asteraceae	*Mikania micrantha*	Nationally invasive
33	Asteraceae	*Parthenium hysterophorus*	Nationally invasive
34	Asteraceae	*Praxelis clematidea*	Nationally invasive
35	Asteraceae	*Solidago canadensis*	Nnationally invasive
36	Asteraceae	*Sphagneticola trilobata*	Nationally invasive
37	Asteraceae	*Synedrella nodiflora*	Nationally invasive
38	Asteraceae	*Tithonia diversifolia*	Nationally invasive
39	Asteraceae	*Tridax procumbens*	Nationally invasive
40	Asteraceae	*Xanthium italicum*	Nationally invasive
41	Asteraceae	*Xanthium spinosum*	Nationally invasive
42	Azollaceae	*Azolla filiculoides*	Nationally invasive
43	Basellaceae	*Anredera cordifolia*	Nationally invasive
44	Brassicaceae	*Lepidium virginicum*	Nationally invasive
45	Cactaceae	*Opuntia dillenii*	Nationally invasive
46	Cactaceae	*Opuntia ficus-indica*	Nationally invasive
47	Cactaceae	*Opuntia monacantha*	Nationally invasive

undefined(continued)

No.	Family	Scientific name	Status
48	Caryophyllaceae	*Agrostemma githago*	Nationally invasive
49	Chenopodiaceae	*Chenopodium ambrosioides*	Nationally invasive
50	Chenopodiaceae	*Chenopodium hybridum*	Nationally invasive
51	Convolvulaceae	*Ipomoea cairica*	nationally invasive
52	Convolvulaceae	*Ipomoea nil*	Nationally invasive
53	Convolvulaceae	*Ipomoea purpurea*	Nationally invasive
54	Convolvulaceae	*Ipomoea triloba*	Nationally invasive
55	Cucurbitaceae	*Sicyos angulatus*	Nationally invasive
56	Euphorbiaceae	*Euphorbia heterophylla*	Nationally invasive
57	Euphorbiaceae	*Ricinus communis*	Nationally invasive
58	Fabaceae	*Leucaena leucocephala*	Nationally invasive
59	Fabaceae	*Mimosa bimucronata*	Nationally invasive
60	Fabaceae	*Mimosa diplotricha*	Nationally invasive
61	Fabaceae	*Mimosa diplotricha* var. *inermis*	Nationally invasive
62	Fabaceae	*Mimosa pudica*	Nationally invasive
63	Fabaceae	*Sesbania cannabina*	Nationally invasive
64	Fabaceae	*Trifolium pratense*	Nationally invasive
65	Fabaceae	*Trifolium repens*	Nationally invasive
66	Geraniaceae	*Geranium carolinianum*	Nationally invasive
67	Labiatae	*Hyptis suaveolens*	Nationally invasive
68	Lythraceae	*Cuphea carthagenensis*	Nationally invasive
69	Malvaceae	*Malvastrum coromandelianum*	Nationally invasive
70	Nyctaginaceae	*Mirabilis jalapa*	Nationally invasive
71	Nymphaeaceae	*Cabomba caroliniana*	Nationally invasive
72	Onagraceae	*Gaura parviflora*	Nationally invasive
73	Onagraceae	*Oenothera biennis*	Nationally invasive
74	Passifloraceae	*Passiflora foetida*	Nationally invasive
75	Phytolaccaceae	*Phytolacca Americana*	Nationally invasive
76	Plantaginaceae	*Plantago virginica*	Nationally invasive
77	Poaceae	*Aegilops triuncialis*	Nationally invasive
78	Poaceae	*Avena fatua*	Nationally invasive
79	Poaceae	*Brachiaria mutica*	Nationally invasive
80	Poaceae	*Bromus catharticus*	Nationally invasive
81	Poaceae	*Cenchrus echinatus*	Nationally invasive
82	Poaceae	*Cenchrus incertus*	Nationally invasive
83	Poaceae	*Lolium temulentum*	Nationally invasive
84	Poaceae	*Melinis repens*	Nationally invasive
85	Poaceae	*Panicum repens*	Nationally invasive
86	Poaceae	*Paspalum conjugatum*	Nationally invasive
87	Poaceae	*Sorghum halepense*	Nationally invasive
88	Poaceae	*Spartina alterniflora*	Nationally invasive
89	Poaceae	*Spartina anglica*	Nationally invasive
90	Pontederiaceae	*Eichhornia crassipes*	Nationally invasive
91	Rubiaceae	*Spermacoce alata*	Nationally invasive
92	Salviniaceae	*Scoparia dulcis*	Nationally invasive
93	Solanaceae	*Datura inoxia*	Nationally invasive
94	Solanaceae	*Datura metel*	Nationally invasive

undefined(continued)

No.	Family	Scientific name	Status
95	Solanaceae	*Datura stramonium*	Nationally invasive
96	Solanaceae	*Solanum aculeatissimum*	Nationally invasive
97	Solanaceae	*Solanum erianthum*	Nationally invasive
98	Solanaceae	*Solanum rostratum*	Nationally invasive
99	Solanaceae	*Solanum torvum*	Nationally invasive
100	Sterculiaceae	*Waltheria indica*	Nationally invasive
101	Umbelliferae	*Daucus carota*	Nationally invasive
102	Verbenaceae	*Lantana camara*	Nationally invasive
103	Verbenaceae	*Stachytarpheta jamaicensis*	nationally invasive
104	Amaranthaceae	*Celosia cristata*	Non-invasive
105	Amaryllidaceae	*Clivia miniata*	Non-invasive
106	Amaryllidaceae	*Hippeastrum vittatum*	Non-invasive
107	Amaryllidaceae	*Narcissus tazetta*	Non-invasive
108	Amaryllidaceae	*Polianthes tuberosa*	Non-invasive
109	Annonaceae	*Annona squamosa*	Non-invasive
110	Apocynaceae	*Nerium indicum*	Non-invasive
111	Araucariaceae	*Araucaria heterophylla*	Non-invasive
112	Arecaceae	*Oenothera odorata*	Non-invasive
113	Arecaceae	*Phoenix dactylifera*	Non-invasive
114	Asteraceae	*Calendula officinalis*	Non-invasive
115	Asteraceae	*Chlorophytum comosum*	Non-invasive
116	Asteraceae	*Gerbera jamesonii*	Non-invasive
117	Asteraceae	*Helianthus annuus*	Non-invasive
118	Asteraceae	*Lactuca sativa*	Non-invasive
119	Asteraceae	*Saussurea costus*	Non-invasive
120	Basellaceae	*Basella rubra*	Non-invasive
121	Bombacaceae	*Bombax ceiba*	Non-invasive
122	Brassicaceae	*Brassica oleracea* var. *botrytis*	Non-invasive
123	Brassicaceae	*Brassica oleracea* var. *capitata*	Non-invasive
124	Brassicaceae	*Camellia sasanqua*	Non-invasive
125	Brassicaceae	*Matthiola incana*	Non-invasive
126	Brassicaceae	*Raphanus sativus*	Non-invasive
127	Cannaceae	*Canna indica*	Non-invasive
128	Capparidaceae	*Cleome spinosa*	Non-invasive
129	Caricaceae	*Carica papaya*	Non-invasive
130	Caryophyllaceae	*Dianthus caryophyllus*	Non-invasive
131	Chenopodiaceae	*Beta vulgaris*	Non-invasive
132	Chenopodiaceae	*Beta vulgaris* var. *cicla*	Non-invasive
133	Chenopodiaceae	*Spinacia oleracea*	Non-invasive
134	Chloranthaceae	*Calathea ornata*	Non-invasive
135	Convolvulaceae	*Ipomoea batas*	Non-invasive
136	Crassulaceae	*Crassula arborescens*	Non-invasive
137	Crassulaceae	*Graptopetalum paraguayense*	Non-invasive
138	Cucurbitaceae	*Citrullus lanatus*	Non-invasive
139	Cucurbitaceae	*Cucumis sativus*	Non-invasive
140	Cucurbitaceae	*Sechium edule*	Non-invasive
141	Cucurbitaceae	*Trichosanthes anguina*	Non-invasive

undefined(continued)

No.	Family	Scientific name	Status
142	Ericaceae	*Begonia semperflorens*	Non-invasive
143	Euphorbiaceae	*Manihot esculenta*	Non-invasive
144	Fabaceae	*Apium graveolens*	Non-invasive
145	Fabaceae	*Arachis hypogaea*	Non-invasive
146	Fabaceae	*Caesalpinia pulcherrima*	Non-invasive
147	Fabaceae	*Cassia fistula*	Non-invasive
148	Fabaceae	*Pachyrhizus erosus*	Non-invasive
149	Fabaceae	*Pisum sativum*	Non-invasive
150	Fabaceae	*Trigonella foenum-graecum*	Non-invasive
151	Fabaceae	*Vicia faba*	Non-invasive
152	Geraniaceae	*Pelargonium domesticum*	Non-invasive
153	Geraniaceae	*Pelargonium graveolens*	Non-invasive
154	Geraniaceae	*Pelargonium hortorum*	Non-invasive
155	Gesneriaceae	*Sinningia speciosa*	Non-invasive
156	Iridaceae	*Crocus sativus*	Non-invasive
157	Iridaceae	*Freesia refracta*	Non-invasive
158	Iridaceae	*Gladiolus gandavensis*	Non-invasive
159	Lamiaceae	*Ocimum basilicum*	Non-invasive
160	Lamiaceae	*Rosmarinus officinalis*	Non-invasive
161	Lamiaceae	*Salvia coccinea*	Non-invasive
162	Liliaceae	*Allium cepa*	Non-invasive
163	Liliaceae	*Allium sativum*	Non-invasive
164	Liliaceae	*Aloe vera*	Non-invasive
165	Liliaceae	*Asparagus setaceus*	Non-invasive
166	Liliaceae	*Tulipa gesneriana*	Non-invasive
167	Magnoliaceae	*Magnolia grandiflora*	Non-invasive
168	Malvaceae	*Gossypium hirsutum*	Non-invasive
169	Malvaceae	*Hibiscus cannabinus*	Non-invasive
170	Moraceae	*Artocarpus heterophyllus*	Non-invasive
171	Moraceae	*Ficus carica*	Non-invasive
172	Musaceae	*Strelitzia reginae*	Non-invasive
173	Myrtaceae	*Psidium guajava*	Non-invasive
174	Myrtaceae	*Syzygium samarangense*	Non-invasive
175	Nolinaceae	*Nolina recurvate*	Non-invasive
176	Nyctaginaceae	*Bougainvillea spectabilis*	non-invasive
177	Nymphaeaceae	*Nymphaea tetragona*	Non-invasive
178	Oleaceae	*Jasminum sambac*	Non-invasive
179	Oleaceae	*Olea europaea*	Non-invasive
180	Onagraceae	*Fuchsia hybrida*	Non-invasive
181	Papaveraceae	*Papaver rhoeas*	Non-invasive
182	Passifloraceae	*Passiflora caerulea*	Non-invasive
183	Platanaceae	*Platanus acerifolia*	Non-invasive
184	Poaceae	*Avena sativa*	Non-invasive
185	Poaceae	*Sorghum vulgare*	Non-invasive
186	Poaceae	*Zea mays*	Non-invasive
187	Polygonaceae	*Fagopyrum esculentum*	Non-invasive
188	Portulacaceae	*Portulaca grandiflora*	Non-invasive

undefined(continued)

No.	Family	Scientific name	Status
189	Primulaceae	*Cyclamen persicum*	Non-invasive
190	Punicaceae	*Punica granatum*	Non-invasive
191	Rosaceae	*Cerasus serrulata* var. *lannesiana*	Non-invasive
192	Rosaceae	*Fragaria ananassa*	Non-invasive
193	Rutaceae	*Ruta graveolens*	Non-invasive
194	Scrophulariaceae	*Antirrhinum majus*	Non-invasive
195	Scrophulariaceae	*Calceolaria crenatifolia*	Non-invasive
196	Solanaceae	*Capsicum annuum* var. *grossum*	Non-invasive
197	Solanaceae	*Capsicum frutescens*	Non-invasive
198	Solanaceae	*Lycopersicon esculentum*	Non-invasive
199	Solanaceae	*Nicotiana tabacum*	Non-invasive
200	Solanaceae	*Petunia hybrida*	Non-invasive
201	Solanaceae	*Solanum mammosum*	Non-invasive
202	Solanaceae	*Solanum tuberosum*	Non-invasive
203	Sterculiaceae	*Pentapetes phoenicea*	Non-invasive
204	Tiliaceae	*Corchorus capsularis*	Non-invasive
205	Tropaeolaceae	*Tropaeolum majus*	Non-invasive
206	Umbelliferae	*Chrysanthemum segetum*	Non-invasive
207	Umbelliferae	*Foeniculum vulgare*	Non-invasive
208	Violaceae	*Viola tricolor*	Non-invasive
209	Vitaceae	*Vitis vinifera*	Non-invasive
210	Zingiberaceae	*Zingiber officinale*	Non-invasive

References

Baker HG (1965) Characteristics and modes of origin of weeds. In: Baker HG, Stebbins GL (eds) The Genetics of colonizing species. Academic Press, New York, pp 147–172

Baker HG (1974) The evolution of weeds. Annu Rev Ecol Syst 5:1–24

Callaway RM, Miao SL, Guo Q (2006) Are trans-pacific invasions the new wave? Biol Invasions 8:1435–1437

Chen L (2012) Effects of environmental factors on plant invasion and risk assessment for invasive species. Sun Yat-sen University, Guangzhou

Chen X (2016) Comparison of photosynthetic characteristics between Asteraceae Invasive. Shenyang Agricultural University, Shenyang

Chen G, Li M, Guo S, Yin L (2009) What attributes correlate with invasiveness of herbaceous angiosperms in Jiangsu, Zhejiang and Shanghai? Acta Ecol Sin 29:5145–5155

Chen L, Peng S, Yang B (2015) Predicting alien herb invasion with machine learning models: biogeographical and life-history traits both matter. Biol Invasions 17:2187–2198. https://doi.org/10.1007/s10530-015-0870-y

Crosti R, Cascone C, Cipollaro S (2010) Use of a weed risk assessment for the Mediterranean region of Central Italy to prevent loss of functionality and biodiversity in agro-ecosystems. Biol Invasions 12:1607–1616. https://doi.org/10.1007/s10530-009-9573-6

Daehler CC (1998) The taxonomic distribution of invasive angiosperm plants: ecological insights and comparison to agricultural weeds. Biol Conserv 84:167–180. https://doi.org/10.1016/S0006-3207(97)00096-7

Daehler CC (2003) Performance comparisons of co-occurring native and alien invasive plants: implications for conservation and restoration. Annu Rev Ecol Evol Syst 34:183–211. https://doi.org/10.1146/annurev.ecolsys.34.011802.132403

Daehler CC, Denslow JS, Ansari S, Kuo HC (2004) A risk-assessment system for screening out invasive pest plants from Hawaii and other Pacific Islands. Conserv Biol 18:360–368. https://doi.org/10.1111/j.1523-1739.2004.00066.x

Dawson W, Burslem DFRP, Hulme PE (2009) The suitability of weed risk assessment as a conservation tool to identify invasive plant threats in East African rainforests. Biol Conserv 142:1018–1024. https://doi.org/10.1016/j.biocon.2009.01.013

Elith J, Graham H, Anderson CP R, et al (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography (Cop) 29:129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Fusco EJ, Finn JT, Balch JK et al (2019) Invasive grasses increase fire occurrence and frequency across US ecoregions. Proc Natl Acad Sci USA. https://doi.org/10.1073/pnas.1908253116

Gassó N, Basnou C, Vilà M (2010) Predicting plant invaders in the Mediterranean through a weed risk assessment system. Biol Invasions 12:463–476. https://doi.org/10.1007/s10530-009-9451-2

González AL, Kominoski JS, Danger M et al (2010) Can ecological stoichiometry help explain patterns of biological invasions? Oikos 119:779–790

Goodwin BJ, McAllister AJ, Fahrig L (1999) Predicting invasiveness of plant species based on biological information. Conserv Biol 13:422–246. https://doi.org/10.1046/J.1523-1739.1999.013002422.X

Gordon D, Onderdonk D, Fox A, Stocker R (2008) Consistent accuracy of the Australian weed risk assessment system across varied geographies. Divers Distrib 14:234–242. https://doi.org/10.1111/j.1472-4642.2007.00460.x

Gordon D, Mitterdorfer B, Pheloung P, Al E (2010) Guidance for addressing the Australian weed risk Assessment questions. Plant Prot Q 25:56–74

Hamilton MA, Murray BR, Cadotte MW et al (2005) Life-history correlates of plant invasiveness at regional and continental scales. Ecol Lett 8:1066–1074. https://doi.org/10.1111/j.1461-0248.2005.00809.x

He J (2012) Exotic plants in China. Shanghai Scientific and Technical Publishers, Shanghai

He S, Yin L, Wen J, Liang Y (2018) A test of the Australian Weed Risk Assessment system in China. Biol Invasions 20:2061–2076

He S, Xu F, Zhang N et al (2020) Applicability of the United States Weed i¼²isk Assessment System to invasive plants in China. Sci Silvae Sin 56:197–208

Hu C, Liu X, Lei Y et al (2016) Foliar nitrogen and phosphorus stoichiometry of alien invasive plants and co-occurring natives in Xishuangbanna. Chin J Plant Ecol 40:1145–1153

Huang G (2014) The biological invasions and quarantine inspection in China. China Quality and Standards Press, Beijing

James JJ, Drenovsky RE (2007) A basis for relative growth rate differences between native and invasive forb seedlings. Rangel Ecol Manag 60:395–400. https://doi.org/10.2111/1551-5028

Kato H, Hata K (2006) Effectiveness of the weed risk assessment system for the Bonin Islands. Assess Control …

Keller RP, Kocev D, Džeroski S (2011) Trait-based risk assessment for invasive species: high performance across diverse taxonomic groups, geographic ranges and machine learning/statistical tools. Divers Distrib 17:451–461. https://doi.org/10.1111/j.1472-4642.2011.00748.x

Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. Trends Ecol Evol 16:199–204. https://doi.org/10.1016/S0169-5347(01)02101-2

Koop AL, Fowler L, Newton LP, Caton BP (2012) Development and validation of a weed screening tool for the United States. Biol Invasions 14:273–294. https://doi.org/10.1007/s10530-011-0061-4

Křivánek M, Pyšek P (2006) Predicting invasions by woody species in a temperate zone: a test of three risk assessment schemes in the Czech Republic (Central Europe). Divers Distrib 12:319–327. https://doi.org/10.1111/j.1366-9516.2006.00249.x

Kueffer C, Pyšek P, Richardson DM (2013) Integrative invasion science: model systems, multi-site studies, focused meta-analysis and invasion syndromes. New Phytol 200:615–633. https://doi.org/10.1111/nph.12415

Lake JC, Leishman MR (2004) Invasion success of exotic plants in natural ecosystems: the role of disturbance, plant attributes and freedom from herbivores. Biol Conserv 117:215–226. https://doi.org/10.1016/S0006-3207(03)00294-5

Levin D (2002) The role of chromosomal change in plant evolution. Oxford University Press, New York

Li X, Zhang M (2015) SPSS 22.0 statistical analysis from Beginner to Master. Publishing House of Electronics Industry, Beijing

Liu GL, Wei SQ (1986) A study of the similarities between China and the world's climate in comparison. J Capital Normal Univ Natural Sci Ed. https://doi.org/10.19789/j.1004-9398.1986.03.010

Lowry E, Rollinson EJ, Laybourn AJ et al (2013) Biological invasions: a field synopsis, systematic review, and database of the literature. Ecol Evol 3:182–196

Ma J (2013) The checklist of the Chinese invasive plants. High reducation press, Beijing

McClay A, Sissons A, Wilson C, Davis S (2010) Evaluation of the Australian weed risk assessment system for the prediction of plant invasiveness in Canada. Biol Invasions 12:4085–4098. https://doi.org/10.1007/s10530-010-9819-3

Morais M, Marchante E, Marchante H (2017) Big troubles are already here: risk assessment protocol shows high risk of many alien plants present in Portugal. J Nat Conserv 35:1–12. https://doi.org/10.1016/j.jnc.2016.11.001

Nishida T, Yamashita N, Asai M et al (2009) Developing a pre-entry weed risk assessment system for use in Japan. Biol Invasions 11:1319–1333. https://doi.org/10.1007/s10530-008-9340-0

Normile D (2004) Expanding trade with China creates ecological backlash. Sci (80-) 306:968–969. https://doi.org/10.1126/science.306.5698.968

Obuchi T, Kabashima Y (2016) Cross validation in lasso and its acceleration. J Stat Mech-Theory E 5:1–37

Pandit MK, Pocock MJO, Kunin WE (2011) Ploidy influences rarity and invasiveness in plants. J Ecol 99:1108–1115. https://doi.org/10.1111/j.1365-2745.2011.01838.x

Pandit MK, White SM, Pocock MJO (2014) The contrasting effects of genome size, chromosome number and ploidy level on plant invasiveness: a global analysis. New Phytol 203:697–703. https://doi.org/10.1111/nph.12799

Pheloung PC, Williams PA, Halloy SR (1999) A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. J Environ Manage 57:239–251. https://doi.org/10.1006/jema.1999.0297

Pyšek P, Richardson DM (2007) Traits Associated with invasiveness in alien plants: where do we stand? In: Nentwig W (ed) Biological Invasion, Ecological studies 193. Springer-Verlag, Berlin, pp 97–125

Pyšek P, Jarošík V, Pergl J et al (2009a) The global invasion success of central European plants is related to distribution characteristics in their native range and species traits. Divers Distrib 15:891–903. https://doi.org/10.1111/j.1472-4642.2009.00602.x

Pyšek P, Krivanek M, Jarosik V (2009b) Planting intensity, residence time, and species traits determine invasion success of alien woody species. Ecology 90:2734–2744

Pyšek P, Manceur AM, Alba C et al (2015) Naturalization of central European plants in North America: species traits, habitats, propagule pressure, residence time. Ecology. https://doi.org/10.1890/14-1005.1

Qiang S (ed) (2009) Weedology. China Agriculture Press, Beijing

Razanajatovo M, Maurel N, Dawson W et al (2016) Plants capable of selfing are more likely to become naturalized. Nat Commun 7:1–9. https://doi.org/10.1038/ncomms13313

Reichard SH, Hamilton CW (1997) Predicting invasions of Woody plants introduced into North America. Conserv Biol 11:193–203. https://doi.org/10.1046/j.1523-1739.1997.95473.x

Rejmánek M (1996) A theory of seed plant invasiveness: the first sketch. Biol Conserv 78:171–181. https://doi.org/10.1016/0006-3207(96)00026-2

Rejmánek M, Richardson DM (1996) What attributes make some plant species more invasive? Ecology 77:1655–1661. https://doi.org/10.2307/2265768

StataCorp (2021) Stata: Release 17. Statistical Software. StataCorp LLC, College Station, TX

Sutherland S (2004) What makes a weed a weed: life history traits of native and exotic plants in the USA. Oecologia 141:24–39. https://doi.org/10.1007/s00442-004-1628-x

Tang NS, Li HQ (2014) Applied regression analysis. Science Press, Beijing

Te Beest M, Le Roux JJ, Richardson DM et al (2012) The more the better? The role of polyploidy in facilitating plant invasions. Ann Bot 109:19–45. https://doi.org/10.1093/aob/mcr277

Thompson K, Hodgson JG, Grime JP, Burke MJW (2001) Plant traits and temporal scale: evidence from a 5-year invasion experiment using native species. J Ecol 89:1054–1060. https://doi.org/10.1046/j.0022-0477.2001.00627.x

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc B 58(1):267–288

Tredennick A, Hooker G, Ellner S, Adler P (2021) A practical guide to selecting models for exploration, inference, and prediction in ecology. Ecology 102:e03336

Van Kleunen M, Dawson W, Schlaepfer D et al (2010) Are invaders different? A conceptual framework of comparative approaches for assessing determinants of invasiveness. Ecol Lett 13:947–958. https://doi.org/10.1111/j.1461-0248.2010.01503.x

Wan F, Guo L, Zhang F (2009) Research on biological invasions in China. Science Press, Beijing

Wang K, Yang J, Chen J (2009) The applications of congeneric comparisons in plant invasion ecology. Biodivers Sci 17:353–361

Weber E, Gut D (2004) Assessing the risk of potentially invasive plant species in central Europe. J Nat Conserv 12:171–179. https://doi.org/10.1016/j.jnc.2004.04.002

Weber E, Sun SG, Li B (2008) Invasive alien plants in China: diversity and ecological insights. Biol Invasions 10:1411–1429. https://doi.org/10.1007/s10530-008-9216-3

Wei Z, Chen X, Yang D, Huang C (2017) Comparison of gas exchange characteristics between two invasive plants and a co-occurring native plant in Liaoning Province. Chin J Eco-Agriculture 25:975–982

Williamson M (1996) Biological invasions. Biol Invasions 334:244. https://doi.org/10.1016/j.crvi.2010.12.008

Williamson MH, Fitter A (1996) The characters of successful invaders. Biol Conserv 78:163–170. https://doi.org/10.1016/0006-3207(96)00025-0

Wu X, Luo J, Chen J, Li B (2006) Spatial patterns of invasive alien plants in China and its relationship with environmental and anthropological factors. J Plant Ecol 30:576–584

Xu H, Qiang S (2004) Inventory of invasive alien species in China. China Environmental Science Press, Beijing

Xu H, Qiang S (2011) China's invasive species. Science Press, Beijing

Yan X, Liu Q, Shou H, Al E (2014) The categorization and analysis on the geographic distribution patterns of Chinese alien invasive plants. Biodivers Sci 22:667–676

Yin L, Yi J, Li W (2007) Weed risk analysis technical requirements

Zhang LJ, Wei XY, Lu JQ, Pan JH (2020) LASSO: from explanation to prediction. Adv Psychol Sci 28(10):1777–1788

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.