



Can Google trends be used to estimate the geographic distribution of alien plants in the United States?

Md Azharul Alam^a, Philip E. Hulme^{a,b,*}

^a Department of Pest-Management and Conservation, Lincoln University, PO Box 84850, Lincoln, Christchurch, New Zealand

^b Bioprotection Aotearoa, Lincoln University, PO Box 84850, Lincoln, Christchurch, New Zealand

ARTICLE INFO

Keywords:

iEcology
Infodemiology
Invasive alien species
Species distribution modelling
Surveillance
Weeds

ABSTRACT

Google Trends has long been used to track the epidemiology of human diseases; however, its application to address biological invasions has been quite limited to date. We develop a workflow for best practice in the use of Google Trends to study biological invasions that accounts for the underlying pitfalls inherent in data from Google searches. We illustrate this workflow by examining the extent Google searches adequately depict the state-wide occurrence of 100 alien plant species in the United States.

Google Trends outputs are based on samples of search queries in the Google search engine during a specific period, and thus, despite many studies using results from a single Google Trends search, we show that it is essential to undertake multiple replicate searches for robust interpretation. In general, results from Google Trends provided only a moderate goodness of fit to the known state-wide occurrence of alien plant species, and then only when the scientific name was used as a keyword. Other keywords, such as the common name or a Topic query, performed poorly. The goodness of fit between the observed species occurrence and that predicted using Google searches was higher for ornamental species or those officially classed by the USDA as noxious or invasive in at least one state. Those states with a greater alien plant richness and higher average education level of their citizens performed best.

Given this data heterogeneity, we highlight an objective workflow to assess the value of Google Trends in order to ensure that greater scrutiny is applied when using this tool in invasion science.

1. Introduction

Plant invasions are a major driver of losses to biodiversity, ecosystem function, and agricultural productivity worldwide (Clements et al., 2022; Foxcroft et al., 2013; Vilà and Hulme, 2017). Efficient management of invasive alien plants depends on effective surveillance and monitoring of their occurrence and spread (Wilson et al., 2017). Traditional methods for detecting and monitoring alien plant species, such as systematic or opportunistic field surveys and remote sensing, are often resource-intensive, time-consuming, slow to capture emerging trends, and geographically patchy (Aschim and Brook, 2019; Nininahazwe et al., 2023; Rakgoale and Ngetar, 2024). Data from citizen scientists are increasingly being viewed as a useful avenue for the surveillance and reporting of alien species distributions because of the high volume of data generated by the public and the opportunity for crowd-sourced species identifications (González-Moreno et al., 2025; Pocock et al., 2024). While citizen science platforms provide an important source of

information to support passive surveillance, the public may also provide data on alien species records through other platforms, such as social media posts (Caley and Cassey, 2023), or by their information-seeking behaviour using search engines or Wikipedia (Mittermeier et al., 2021).

The concept of iEcology highlights that human activity on the internet can provide important ecological information about species distributions (Jarić et al., 2020; Jarić et al., 2021; Novoa et al., 2025). By capturing changes in search activity over time and in different locations, these tools may offer indirect indicators of species distributions, conservation awareness, or concern (Jarić et al., 2020). The application of iEcology builds on the long history of infodemiology and infoveillance in which web-based information has been used in the analysis, detection, and forecasting of human diseases and epidemics (Mackey et al., 2022; Mavragani, 2020). Ecologists have used web-based information from various online sources, such as posts and images on Facebook and Twitter/X (Daume, 2016; Marcenò et al., 2021), videos on YouTube and other platforms (Dylewski et al., 2017; Jagiello et al., 2019; Miranda

* Corresponding author at: Department of Pest-Management and Conservation, Lincoln University, PO Box 84850, Lincoln, Christchurch, New Zealand.

E-mail address: Philip.Hulme@lincoln.ac.nz (P.E. Hulme).

<https://doi.org/10.1016/j.ecoinf.2026.103689>

Received 10 November 2025; Received in revised form 1 March 2026; Accepted 1 March 2026

Available online 3 March 2026

1574-9541/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

et al., 2016), and images on Flickr and Google Images (Allain, 2019; ElQadi et al., 2017; Leighton et al., 2016) to map and monitor species occurrences across the globe. Another such platform is Google Trends, which is a free, easily accessible, and publicly available tool that summarises the search volume for specific keywords entered into the Google search engine at a national, regional, or city level (Cebrián and Domenech, 2023).

Although Google Trends has been used as a tool for the infodemiology and infoveillance of human diseases for more than a decade (Pervaiz et al., 2012; Seifter et al., 2010), it has only recently caught the attention of invasion scientists. Nevertheless, despite this short history, Google Trends has proven to be useful in investigating public attention and awareness towards various alien species. In general, Google search volume appears to be correlated with the spatial distribution of alien plants, insects, and fish within a region (Fazzari et al., 2024; Fukano and Soga, 2019; Maciejewski et al., 2025; Proulx et al., 2014; Woodworth et al., 2023), suggesting that data from Google Trends could be a suitable tool for surveillance and early warning (Jarić et al., 2020; Novoa et al., 2025). However, despite the rising popularity of using Google Trends to examine the spatial patterns of biological invasions, much of the research to date has paid scant attention to three major pitfalls of using this tool.

First, Google Trends allows users to search for a species using a specific search term such as its scientific name or common name but also provides the option for a Topic search. Topic searches capture the broader semantic meaning, including synonyms, different languages and related concepts, potentially offering a comprehensive view of interest in a species, while a standard search focuses only on the exact query and misses variations. The implications of using different search keywords are unknown, and while Topic searches are frequently chosen due to their often greater search volume (Fazzari et al., 2024; Fukano and Soga, 2019), they can generate spurious results by including keywords unrelated to the term of interest (Rovetta, 2024). Second, each time a search is made using Google Trends, the resulting output is only based on a sample of Google searches, and this can lead to considerable variation in the output when keywords have low overall search volumes and for regions with low internet penetration rates or low human population size (Hölzl et al., 2025). Third, Google Trends normalises results relative to the total number of Google searches made during the same time window, but search volume varies over both space and time, which, unless these biases are controlled for, makes it an unreliable metric for spatiotemporal studies of biological invasions (Cebrián and Domenech, 2024; Lolić et al., 2024). Fourth, analyses at a global scale need to recognize the varying popularity of different search engines in individual countries or regions (Springer et al., 2025). While Google dominates globally with almost 90% of the search engine market, Baidu is the most popular search engine in China, and this is also the case for Yandex in Russia and Naver in South Korea (Polak, 2025). These four concerns have been recognized only relatively recently within the infodemiological literature and while some solutions have been proposed (Cebrián and Domenech, 2024; Rovetta, 2024; Springer et al., 2025) there is as yet no universally agreed workflow.

We propose a standardized workflow for examining the utility of Google Trends to adequately map the distribution of multiple alien plant species. Specifically, we explore whether search interest using Google can be used to estimate the state-wide occurrence of 100 alien plant species in the United States. First, we explore the significance of the choice of keyword in determining the accuracy in the state-wide occurrence of a species and the importance of including search data from multiple Google Trends queries. Second, we also investigate the ecological attributes of individual alien plants and the socioeconomic attributes of individual states that make Google searches more reliable indicators of the actual state-wide occurrence of alien plant species. Having a robust workflow for using Google Trends to map alien plant species occurrence accurately should significantly enhance monitoring efforts and enable the prioritization of management strategies (Novoa

et al., 2025).

2. Materials and methods

We focused our study on the state-wide occurrence of alien plant species found in the United States. The United States was chosen as the study area because Google is the dominant search engine and it has a consistent search language in English for common names, in addition the individual states have sufficient population sizes (500,000+ people), as well as a high internet penetration rate (> 90%) to ensure reliable search volume. Subregional data from Google Trends have been widely used in other disciplines to assess state-wide interest in diverse topics, including child abuse (Hartwell et al., 2022), depression (Wang et al., 2022), and tick-borne disease risk (Yang et al., 2024). The alien plant species included in our study were selected with regard to their life form and their commonness across the United States. The alien plant species data were sourced from an online database of 1548 plant species identified as invasive in natural areas, published by the Invasive Plant Atlas of the United States (www.invasiveplantatlas.org/distribution.cfm). To control for any biases in search effort related to plant characteristics, 20 species were selected from each of the following five life-forms: forbs, grasses, shrubs, trees, and vines (Table S1). To avoid problems with fitting models where species are either ubiquitous or particularly rare, only those species that occurred in 10 to 40 states were included to ensure robustness of statistical models.

2.1. National species occurrence records

For each of our target 100 alien plant species (Table S1), we collected their most recent state-wide occurrence data from the USDA PLANTS database (<https://plants.usda.gov/> accessed 17/03/25). Additionally, we gathered public search data for each of the 100 invasive alien plant species using Google Trends (Google LLC, Mountain View, California, USA; <https://trends.google.com/trends/>). We focused solely on web searches but included all search categories since this generally gives the best results for different queries. For each species, we used the currently agreed scientific name and common name as provided on the USDA Plants database as literal keywords (exact matches) and included each term in quotation marks. In addition, Google Trends often provides broader conceptual keywords (semantic matches) that include additional terms that Google associates with either the scientific or common name (Mavragani and Ochoa, 2019). Google Trends automatically suggests alternative terms when a specific search term is entered. For example, entering the specific term “*Euonymus europaeus*” prompts Google Trends to offer ‘European spindle’ as a Plant search. Conversely, entering the specific common name ‘European spindle’ provides the option to search for ‘European Spindle’ as a Topic (note the difference in capitalization). We therefore included two conceptual keywords described as “Plant” and “Topic” keywords, with the main difference between the two being that the former recognized that the term was a plant (all search terms are listed in Table S1). While conceptual keywords may pick up variations in common names or synonyms of scientific names, they also run the risk of confounding closely related species and even items completely unrelated to plants. All searches were undertaken manually rather than through an automated API search.

Since we aimed to assess whether public interest in a species, as captured by Google searches, reflected the actual occurrence of a particular species in a US state, we included data on all public searches undertaken in the entire United States and retrieved “Interest by sub-region” data over a twenty-one-year period, from January 1, 2004, to December 31, 2024. The long period of time over which data were collated had the benefit of capturing a large volume of public search data that avoided any biases relating to temporal changes in public search volumes or interests (Lolić et al., 2024).

For each species individually, we examined Google Trends data for each of the four keywords as separate queries and recorded in which of

the states a relevant public search had occurred for that literal or conceptual keyword. Due to the volume of data generated by public searches using Google, the output from Google Trends is only based on a sample of the entire data set, and thus queries on different days or from different IP addresses generate different outputs (Lolić et al., 2024). We therefore undertook the queries for all 100 species and the four keywords on a single day and replicated this approach across ten separate days to capture a different sample of search data. For each search, Google Trends returns a measure of the Relative Search Volume (RSV), a value from 0 to 100. To avoid the pitfalls of overinterpreting variation in RSV as a meaningful indicator of absolute interest (Liu, 2024), we converted the Google Trends output into a binary variable. Whenever there was evidence of the keyword being searched using Google in a US state, we recorded the value as 1, whereas when Google Trends returned no such evidence, the value was recorded as 0. We included values returned by Google Trends as <1 as evidence of a search for the keyword. For the final analysis, we calculated the average of the scores from the ten independent samples, resulting in a single continuous predictor variable for each state and species ranging from 0 (species not recorded from any searches) to 1 (species recorded in all ten searches).

2.2. Species attributes that might influence search intensity

We expected species that were known to be invasive, used as ornamentals, or had a shrub or tree lifeform to be more familiar to the general public, and thus Google searches might better reflect their actual occurrence in a state. We classified the species according to whether they were recorded in the USDA PLANTS database as invasive, noxious, prohibited, or restricted (<https://plants.usda.gov/noxious-invasive-search/> accessed on 26/05/2025). We subsequently categorized the species into two invasiveness categories: “high,” indicating that the species was listed as invasive or noxious in at least one state, and “low,” indicating that the species was not listed in any state as invasive or noxious. In total, 52 species were categorized as low, while 48 were categorized as high (Table S1). We classified the ornamental status of each species in relation to whether it was recorded as such in either the CABI Invasive Species Compendium database (<https://www.cabidigitallibrary.org/product/QI/> accessed on 26/05/2025) or the Encyclopedia of Plants and Flowers (Brickell, 2011). In total, 63 species were categorized as ornamental (Table S1).

2.3. State-wide variables that might influence search intensity

We expected Google search intensity (as a proxy for interest in alien plant species) to vary among different states as a function of attributes such as human population size, average level of education, and number of alien plant species in the state. Thus, for each state, we extracted variables that might explain the frequency with which species occurrence had been correctly predicted in each state using Google searches. We focused on five relatively uncorrelated variables ($-0.6 < r < 0.6$; $VIF < 2$; Fig. S1) for which we had clear expectations (Table 1).

2.4. Data analysis

To evaluate how well public interest, as measured by Google searches, could predict the actual state-wide occurrence of each alien plant species, we used the roc() function to generate Receiver Operating Characteristic (ROC) curves and calculated the Area Under the Curve (AUC), indicating goodness of fit for each keyword used in our Google Trends queries. Analyses were undertaken separately for each species. In each case, our binary independent variable was the actual species state-wide occurrence (presence or absence) in the United States, and our explanatory variable was the average score of the ten Google Trends samples that ranged from 0 (species not recorded from any searches) to 1 (species recorded in all ten searches). This provided us with an AUC score for each species based on how well a particular keyword search on

Table 1

The state-wide socioeconomic variables, their rationale for inclusion, and the sources from which they were collected.

Variables	Rationale	Source
Number of museums	Increase public awareness and interest in alien plants through exhibitions, workshops, and outreach programs, which can positively drive public search activity	Wikipedia (https://en.wikipedia.org/wiki/List_of_museums_in_the_United_States accessed on June 16, 2025)
Proportion of the population with advanced education	With a greater average level of education in a state, citizens are likely to have greater knowledge and awareness of alien species and their impacts on ecosystems, which can increase public search activity	United States Census Bureau (https://www.census.gov/en.html accessed on June 16, 2025)
Proportion of the population using the internet	More people with access to the internet result in more searches and thus likely to have higher search volumes for alien plant species	National Centre for Education Statistics (NCES; https://nces.ed.gov/ accessed on June 16, 2025)
Number of botanic gardens	Botanic gardens often introduce and showcase varieties of alien plant species, which can increase public awareness, interest, and search activity	BGCI garden search (https://gardensearch.bgci.org/ accessed on June 16, 2025)
Alien species richness	Having more alien plants in a state can increase public awareness and concern due to their negative impacts, which increases the public search activity	The total number of alien plants from our selected list of 100 species that occur in the state

Google captured the actual state-wide species occurrence.

To examine differences in the goodness of fit of each of the four Google Trends keyword types (scientific name, common name, Plant query, and Topic query), we performed a one-way repeated measures ANOVA. The AUC scores were the dependent variable, and the four keywords were treated as different levels of a single factor, while species was modelled as “repeated” within each level. Since the assumption of sphericity was violated, the Greenhouse-Geisser sphericity correction was automatically applied. The significant differences in means were found using Tukey’s HSD tests. We compared the frequency with which each keyword resulted in false positives and false negatives regarding the actual state-wide species occurrence using a Kruskal-Wallis test, followed by Dunn’s post-hoc test.

To investigate the influence of species-level attributes on AUC scores based on the best-performing keyword, we examined the variation across the five growth forms using a Kruskal-Wallis test, while Wilcoxon rank-sum tests examined differences in relation to binary predictors: woodiness (woody vs non-woody), invasiveness (high vs low), and ornamental status (yes vs no). We checked data normality between groups by performing the Shapiro-Wilk test and the equality of variances between groups by performing Bartlett’s test. We then evaluated the relative importance of these predictors in explaining variation in AUC scores by fitting a linear model and then used the Akaike Information Criterion (AIC) for model comparison, assessing the impact of excluding each predictor from the full model. The response variable was logit-transformed to fulfil the data normality assumption. We also calculated the relative contribution of each variable to the model following the lmg method. The strength of association between these factors was quantified using Cramér’s V.

To investigate the factors that determined variation in the state-wide mean AUC score based on the best-performing keyword, we ran a stepwise regression analysis (in both directions) and selected the best

model based on the AIC scores. Starting from the null model (intercept-only), predictors were added (forward selection) and then removed (backward elimination step) based on improvements in AIC. The relative importance of each variable was also calculated following the lmg method.

All the statistical tests were performed in R (version 4.3.2) (R Core Team, 2023) using the packages pROC (Robin et al., 2011), tidyverse (Wickham et al., 2019), ggpubr (Kassambara, 2020), rstatix (Kassambara, 2021), patchwork (Pedersen, 2019), agricolae (de Mendiburu, 2019), psych (Revelle, 2019), corrplot (Wei and Simko, 2024), mass (Ripley, 2011), car (Fox and Weisberg, 2018), relaimpo (Grömping, 2007), vcd (Meyer et al., 2020), GGally (Schloerke et al., 2021) and usmap (Di Lorenzo, 2025).

3. Results

We extracted Google Trends data for all 100 species for the scientific name and common name queries, for 88 species using Plant queries, and for 67 species using Topic queries. Due to the nature of Google Trends samples, it is expected that the adequacy of searches reflecting the actual state-wide species occurrence will increase with a greater number of independent samples. We assessed the adequacy of using ten independent queries by examining, for each species, how the cumulative number of US states with detectable Google search activity increased as additional samples were added. We assumed that this cumulative count would stabilize after only a few queries. However, we identified three distinct patterns showing that, in many cases, the number of states in which a species was searched increased with the number of independent Google Trends queries (Fig. 1). While a small number of species for scientific name (2 species), common name (20 species), Plant query (27 species), and Topic query (7 species) showed no change in the cumulative number of states with increasing number of samples, most showed a rapid rise in the number of states to an asymptote: scientific name (69 species), common name (55 species), Plant query (55 species) and Topic query (57 species). Nevertheless, several species still had not reached an asymptote even after ten independent samples: scientific name (29 species), common name (25 species), Plant query (6 species), and Topic query (3 species). When using either the scientific name as a keyword or a Topic query, there was no indication that the AUC score differed among species that showed an asymptotic, increasing, or stable search trend (Fig. S2). In contrast, for both the common name keyword and Plant queries, the stable pattern showed a significantly lower AUC score (Fig. S2).

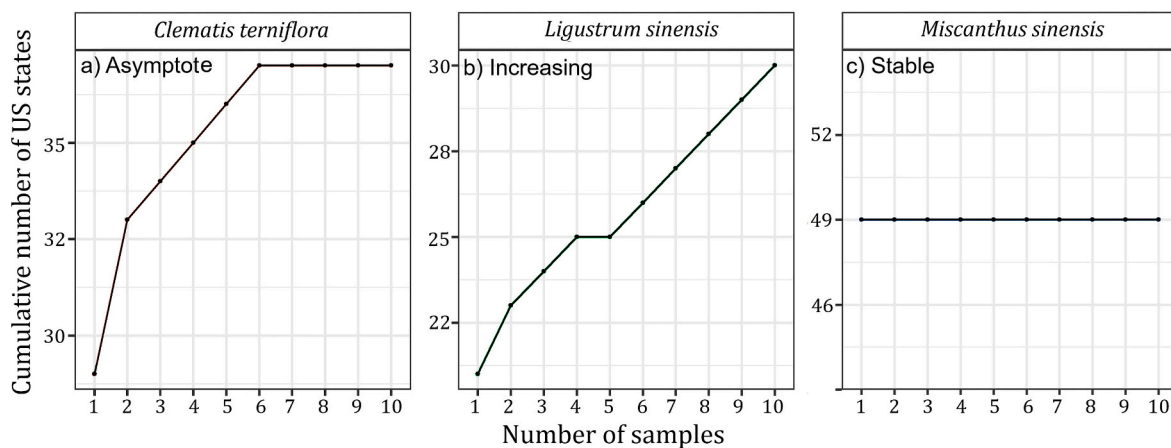


Fig. 1. Examples of how the cumulative number of US states for which there was evidence of a Google search changed over ten independent queries using Google Trends for the scientific name of three alien plant species: a) Asymptote, indicating a rapid rise in the number of states and then reaching a plateau of no or limited further change with more samples; b) Increasing, indicating the continuous accumulation of states with additional samples; c) Stable, indicating no changes in number of states with increased number of samples.

3.1. Accuracy of different keywords

On average, the predictive ability of Google Trends keywords was at best fair and, at worst, no better than random (De Hond et al., 2022). Across the four keywords, the highest AUC was observed for the scientific name (0.66 ± 0.01), followed by Plant queries (0.61 ± 0.01), and common names (0.60 ± 0.01), while Topic queries consistently demonstrated the lowest predictive performance (0.51 ± 0.01) (Fig. 2). As a result, the four keywords significantly differed in their ability to predict actual species state-wide occurrence ($F_{(3, 162)} = 40.76$, $p < 0.001$, generalized eta squared (η^2G) = 0.348). The goodness of fit between the actual occurrence of a species and that predicted from Google searches depended on the keyword used. For over 72% of the species, the scientific name had the highest AUC score, but for 28%, it was the common name. Overall, the rank order of species in terms of their AUC score was not correlated between the four different keywords, highlighting that not only did different search terms provide different goodness of fit, but also that these patterns depended on the species in question.

Data from Google Trends tended to generate more false negatives than false positives for both the scientific name and common name, highlighting that these keywords often failed to predict the true presence of a species in a state (Fig. 3). This trend was also true for Topic queries, but to a much greater extent. In contrast, the low goodness of fit of Plant queries was driven by a high number of false positives, emphasising that this broader search term tended to overestimate the distribution of alien plants in the United States (Fig. 3).

3.2. Correlates of species-level variation in goodness of fit

Even when using the scientific name as the keyword, AUC scores still varied widely among alien plant species, with scores ranging from 0.45 to 0.99, indicating that while the state-wide distribution of some alien plant species was well predicted, other predictions were no better than random (Fig. 4, Fig. S3). Two species, *Triadica sebifera* and *Microstegium vimineum*, demonstrated excellent goodness of fit with AUC scores greater than 0.90. Additionally, 31 species had AUC scores ranging from 0.70 to 0.90, indicating reasonable predictive performance; among these were *Melia azedarach*, *Perilla frutescens*, and *Lonicera maackii*. In contrast, 34 species had AUC scores below 0.60, suggesting poor predictions for these species.

When using the scientific name as the keyword, we did not find any significant differences in AUC scores across the five growth form categories ($\chi^2 = 8.17$, $p = 0.086$, Fig. 5a). However, when we categorized the

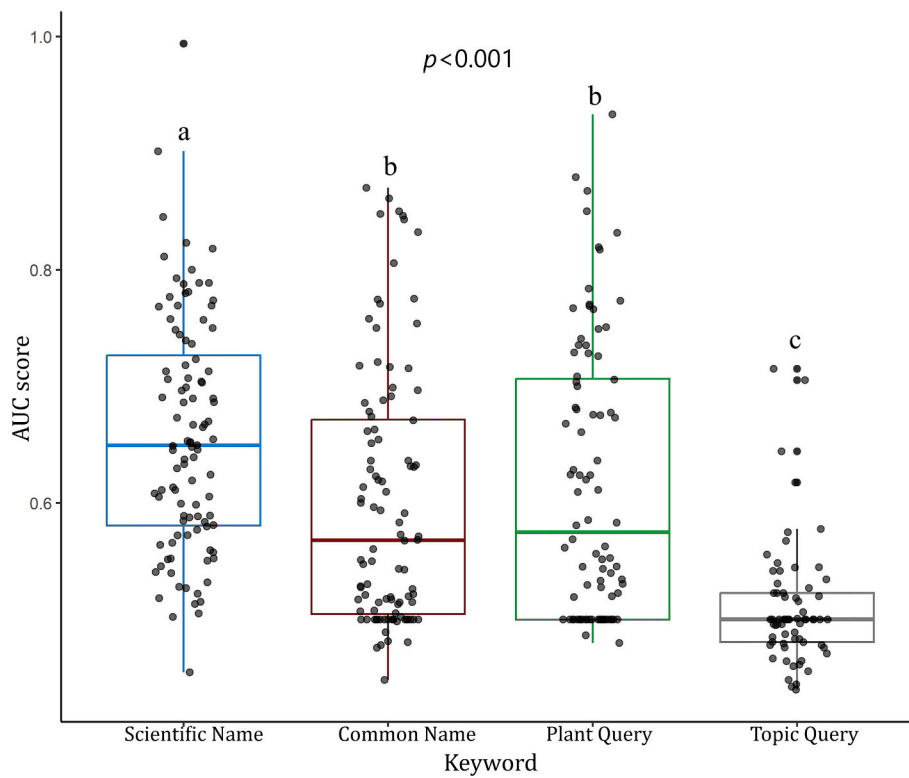


Fig. 2. Average goodness of fit (as determined by the AUC score) between the actual occurrence of a species and that predicted from Google searches for each of four different keywords. Boxplots show the median, interquartile range, and range of AUC scores. Groups with different letters differ significantly from each other.

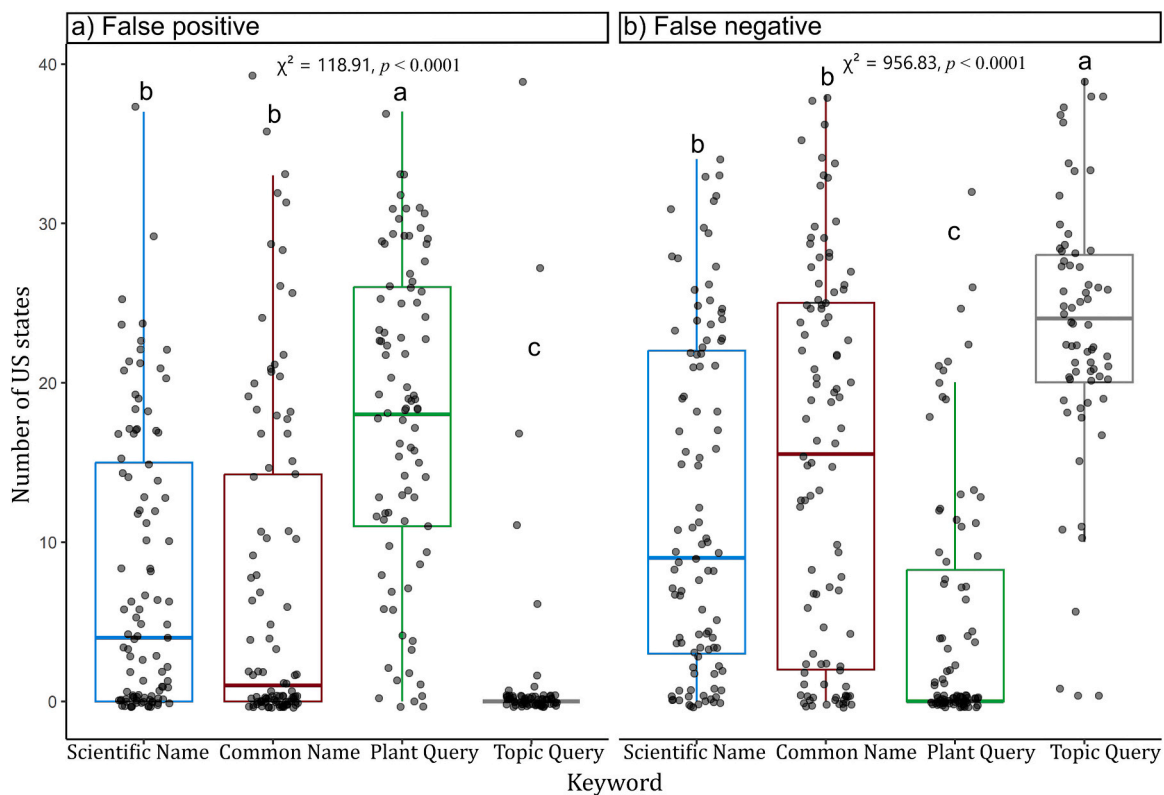


Fig. 3. Variation in the number of states recorded as a) false positives (absent from a state but Google searches suggest presence) and b) false negatives (present in a state but Google searches suggest absence) for alien plant species searched using four keywords. Boxplots show the median, interquartile range, and range of state numbers. Groups with different letters differ significantly from each other.

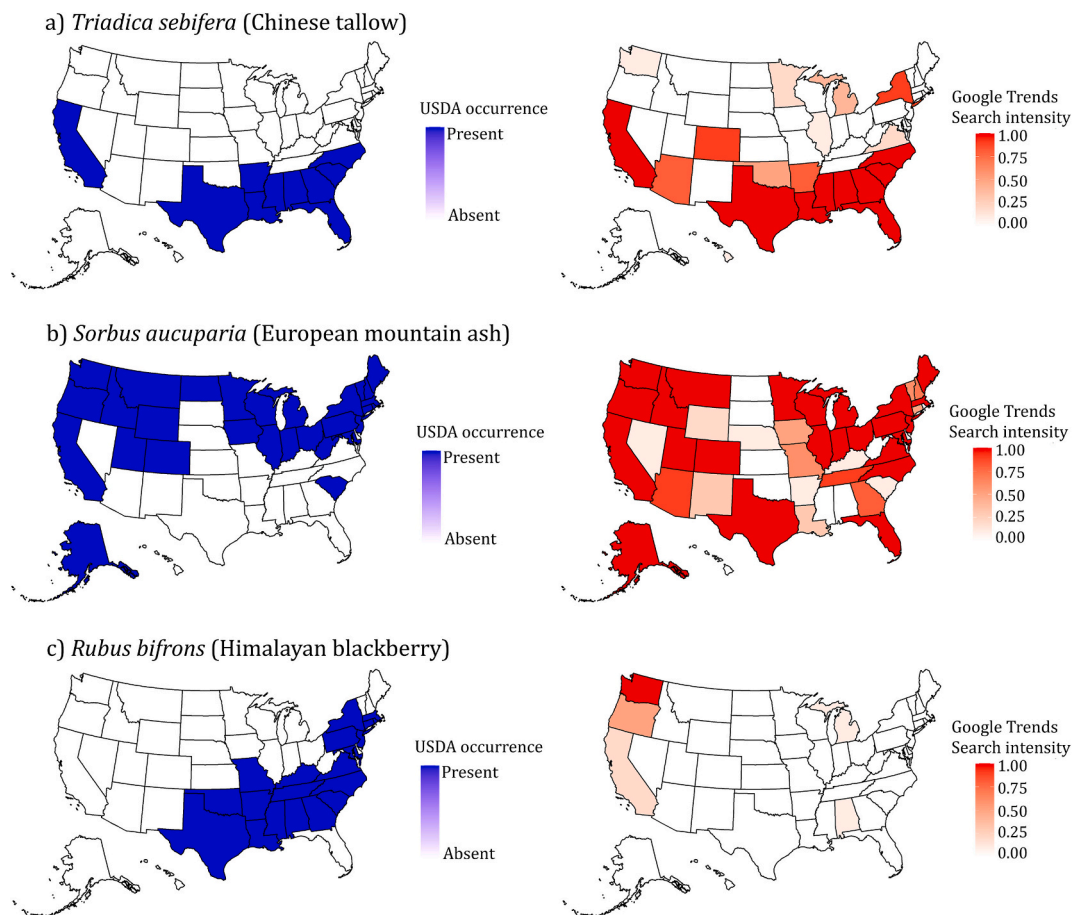


Fig. 4. Maps visualizing the USDA occurrence (left panel) and the average score from ten Google Trends queries based on using the scientific name as a keyword (right panel) of a) a well-estimated species (*Triadica sebifera*: AUC = 0.99), b) a moderately well-estimated species (*Sorbus aucuparia*: AUC = 0.75), and c) a poorly estimated species (*Rubus bifrons*: AUC = 0.45).

species based on their woodiness, we found a statistically significant difference in AUC scores between the two groups ($W = 883$, $p = 0.013$), indicating that woody species had better goodness of fit compared to non-woody species (Fig. 5b). Additionally, we found that ornamental species had higher AUC scores than non-ornamental species ($W = 540.5$, $p < 0.0001$, Fig. 5c), and highly invasive species showed higher AUC scores than those characterized by low invasive potential ($W = 1725.5$, $p < 0.001$, Fig. 5d).

The model, including woodiness, invasiveness, and ornamental status, showed that invasiveness and ornamental status were the only significant predictors retained ($p < 0.001$) in the model (Fig. S4). The AIC-based model comparison further determined that excluding woodiness even slightly improved the model fit ($\Delta AIC = -2.0$) (Fig. S4). The woodiness status was strongly associated with the ornamental status ($V = 0.50$, $p < 0.0001$), and the relative importance analysis further confirmed that woodiness status contributed negligibly to the explained variance.

3.3. Correlates of state-wide variation in goodness of fit

The AUC score for each state (estimated over all 100 species) also varied widely across different states but was generally poor, with scores ranging from 0.45 to 0.71 (Fig. 6). Pennsylvania (0.71) and Ohio (0.68) had the highest mean AUC scores across all species, suggesting Google search interest more accurately reflects the actual occurrence of alien plant species. In contrast, states such as Montana (0.45), Texas (0.46), and Nebraska (0.47) had particularly low mean AUC scores across all 100 alien plant species.

We found that state-wide AUC scores were significantly correlated with how many of our 100 alien species were found in that state (alien species richness) and the percentage of the population with advanced education (Fig. S1). The stepwise AIC model selection identified alien species richness as the strongest predictor of AUC ($AIC = -284.93$, $R^2 = 0.18$, $p = 0.0017$). Relative importance analysis further showed that alien species richness and percentage of population with advanced education highly contributed to the explained variance, with alien species richness being the dominant predictor, explaining $\sim 62\%$ and percentage of population with advanced education explaining $\sim 26\%$ of the model's explanatory power (Table 2). Both variables had a positive relationship with state-wide AUC scores, indicating that states with greater alien species richness and a larger population with advanced education had a higher goodness of fit (Fig. S5).

4. Discussion

Despite the promise of iEcology to provide new insights into the dynamics of alien species, evidence supporting the utility of internet-based search activity to capture facets of biological invasions has to date been limited (Jarić et al., 2021; Novoa et al., 2025). We aimed to develop a standardized workflow to effectively evaluate the potential of Google searches to capture the geographical distribution of alien plant species in the United States (Fig. 7). Previous studies have used search volumes extracted from Google Trends to quantify the spatiotemporal dynamics in public interest of alien species (Fazzari et al., 2024; Fukano and Soga, 2019; Woodworth et al., 2023) or to map their geographic distribution (Proulx et al., 2014). However, these studies relied on

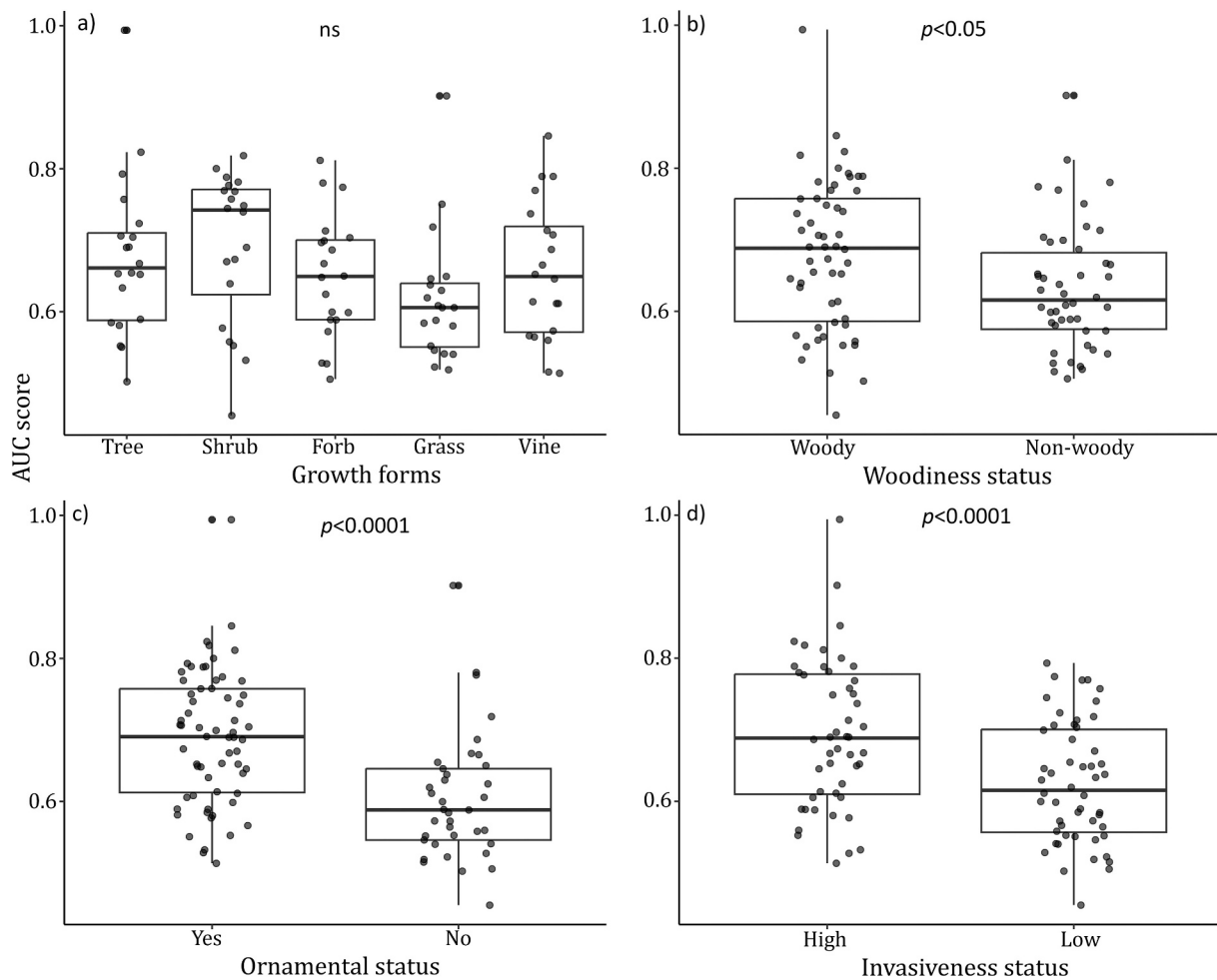


Fig. 5. Comparison of goodness of fit (as determined by the AUC score) between the actual occurrence of an alien plant species and that predicted from Google searches in relation to a) growth forms, b) woodiness status, c) ornamental status, and d) invasiveness status in the United States. Boxplots show the median, interquartile range, and range of AUC scores.

single-query and single-keyword methods to extract information, hence suffer from limitations that can result in misleading or inaccurate interpretations (Cebrián and Domenech, 2024; Hölzl et al., 2025; Lolić et al., 2024; Rovetta, 2024). We addressed these issues by extracting Google Trends data from ten independent queries across four different keywords over multiple different species, which reduces sampling bias and improves the robustness of geographical pattern detection. Our research highlights the potential of using multiple independent queries and averaging results to minimize the random noise associated with single queries and thus generate more reliable and reproducible search data (Cebrián and Domenech, 2024). Even so, we highlight that even with 10 samples, the number of states for which species was recorded had not reached a plateau in some cases. Nevertheless, we present a robust approach to assess the value of using the Google searches as a tool to describe alien species occurrence, identify the importance of selecting appropriate keywords as search terms, and assess objectively under what circumstances Google searches may be valuable in the management of biological invasions.

Although common names have previously been used to assess the value of Google searches in capturing the number of records of alien plants across the United States (Woodworth et al., 2023), we found that, on average, using the scientific name of a plant species was a more effective keyword than using the common name, or, for that matter, Plant or Topic queries. There are several reasons why common names might be less likely to reflect the occurrence of a species in a particular state. First, common names may vary even within a state, e.g., *Schinus*

terebinthifolius is known in Florida as Brazilian pepper, Brazilian peppertree, Florida holly, and Christmas berry. Second, the same species may have different common names across the United States, e.g., *Leucaena leucocephala* is known as white lead tree in California but koa haole in Hawai'i. Third, there are different species that share the same common name in the United States, such as ironwood, which is used for both the alien *Casuarina equisetifolia* as well as the native *Ostrya virginiana*. Furthermore, it might be expected that individuals using the scientific name as a search term may have greater knowledge regarding the species that is the subject of their search than those using the common name.

Scientific names are not without their challenges either due to problems of synonymy. For example, *Reynoutria japonica* is the accepted name for Japanese knotweed, but this scientific name is much less frequently used as a Google search term than either the synonyms *Polygonum cuspidatum* or *Fallopia japonica*. Indeed, the USDA Plants database favours the synonym *P. cuspidatum*, which is also the most frequently used Google search term for Japanese knotweed. Woodworth et al. (2023) used common names in their searches for their selection of species, as not all scientific names presented a sufficiently high search volume for their analyses. This might have arisen because common synonyms as used in the USDA Plants database were not examined as possible alternative search terms to the widely accepted names, as we found no difficulty in obtaining adequate search volumes using the USDA Plants scientific names. Topic and Plant queries should, at least in theory, avoid such problems since these searches include scientific and

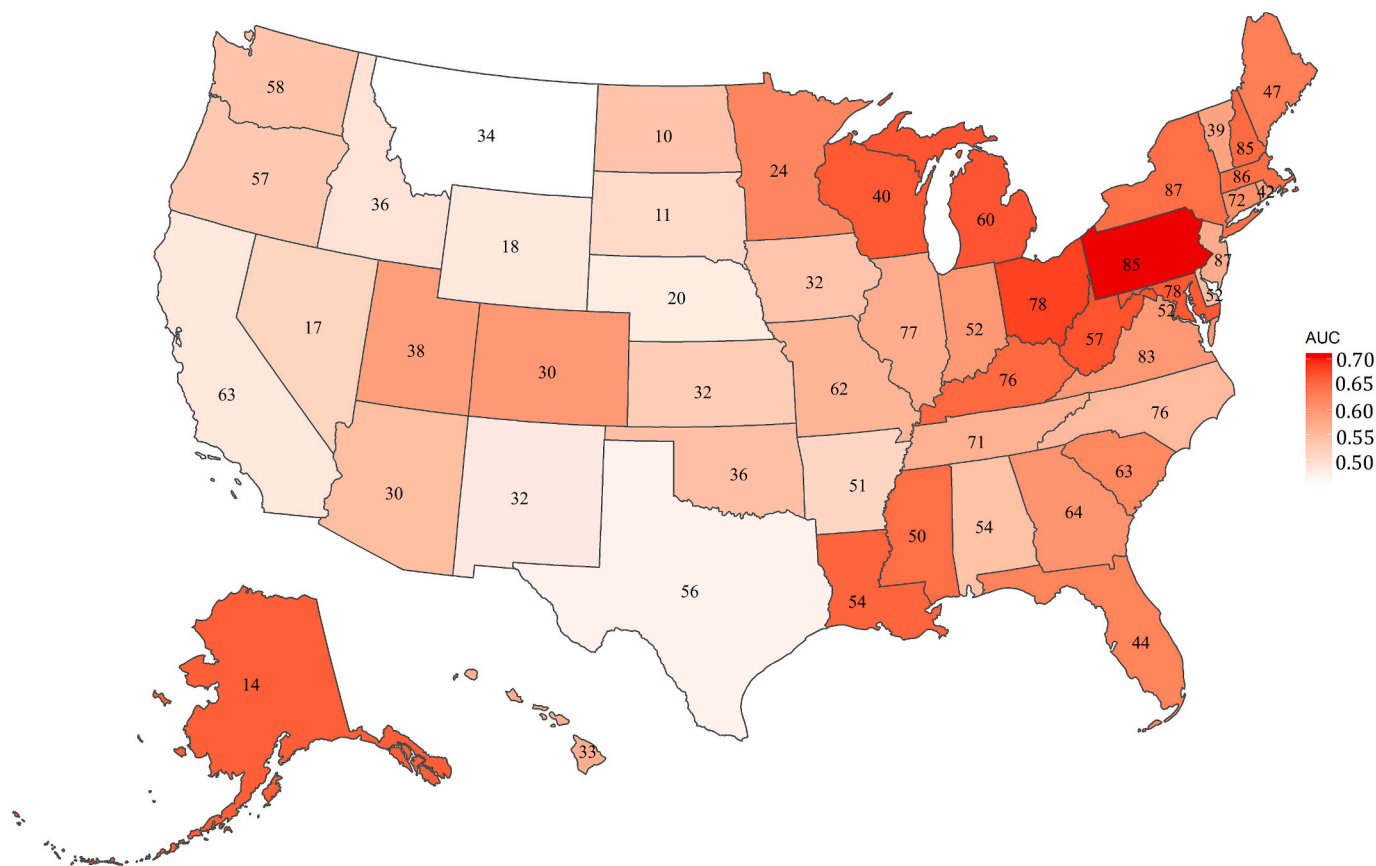


Fig. 6. Map visualizing the predicted distribution of alien plant species across different states in the United States, based on Google searches using the scientific name as a keyword. Each state is shaded from white to red, indicating the AUC score for that state that reflected the ability of Google searches to accurately estimate the occurrence of individual species in the state. The number displayed within each state represents the total number of alien plant species out of the 100 we examined that had been recorded as present in that state. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2

AIC-based stepwise regression model outputs showing variables with their AIC, ΔAIC, and variable importance score.

Variable	AIC	ΔAIC	Variable importance
Alien species richness	-284.93	0.00	0.617
Proportion of population with advanced education	-284.66	0.27	0.264
Number of botanic gardens	-283.52	1.41	0.059
Proportion of the population using the internet	-283.25	1.68	0.007
Number of Museums	-283.05	1.88	0.052

common names as well as synonyms, which can result in a higher search volume. However, they are likely to include different unrelated information (Rovetta, 2024). Our results demonstrated that Topic queries performed the worst of all search terms, with evidence for very few Google searches for most of the species, highlighting the relative inefficacy of Topic queries in capturing public interest and awareness regarding alien plant species. This may also explain the propensity of Topic queries to produce many false negatives. Conversely, Plant queries picked up more Google searches, suggesting that the algorithm used to generate Plant searches might have included too many unrelated terms, making it less reliable for accurately reflecting the actual state-wide occurrence of specific alien plant species. For this reason, Plant queries tended to generate too many false positives. Additionally, both the Topic query and Plant query are not available for all species on Google Trends, which limits their utility in investigating certain species.

We demonstrated only a fair connection between Google search intensity using the scientific name and the actual state-wide occurrence of alien plant species across US states. The state-wide distribution of more than half of our alien plant species was moderately to well predicted by Google searches. This shows that public search interest can capture, to some extent, large-scale species occurrences in the real world, but probably not at the level of accuracy required for surveillance or management. We found that the goodness of fit between Google searches and the state-wide occurrence of a species depended on the ecological characteristics of the plant, suggesting that alien plant species with certain traits are better suited to iEcology studies. There was comparatively a better goodness of fit between the actual occurrence of an alien plant species and that predicted from Google searches for plant species with ornamental use and high invasiveness potential. Ornamental species feature distinct visual characteristics, such as colourful flowers and leaves or unique shapes and architecture, making them easily recognizable and increasing public interest, which in turn boosts the frequency with which they are likely to be searched (Altman et al., 2022; Andrade et al., 2025; Vardi et al., 2021). Invasive alien species often attract more public attention due to their adverse effects on agriculture, public health, and the economy, as well as their dominance in news and awareness campaigns highlighting extensive management efforts, resulting in an increased number of searches (Fukano and Soga, 2019; Wong, 2024; Woodworth et al., 2023).

While we demonstrated that for some alien species, Google searches can capture their actual state-wide occurrence, it is essential to recognize that public interest in a plant species can be influenced by various other factors, and therefore, may distort its utility in understanding

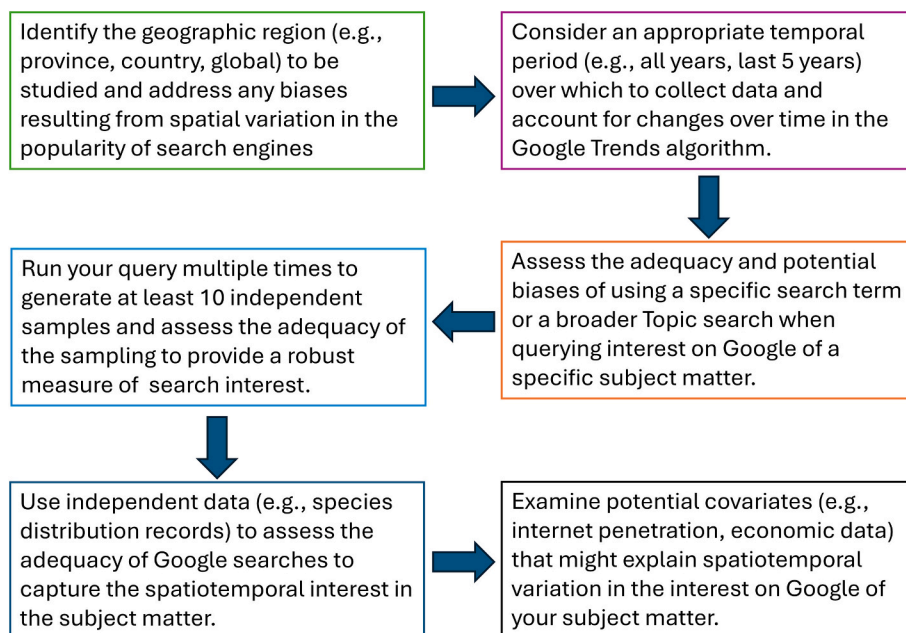


Fig. 7. Schematic workflow of the issues that need to be considered when using Google Trends to investigate interest in a particular subject matter. The workflow has been designed to be generic for any subject matter though it has been developed from an examination of the ability of Google searches to capture the state-wide distribution of alien plants in the United States.

ecological patterns (Woodworth et al., 2023). Social media, news reporting on television and radio, public awareness campaigns, or particular health risks of a plant can specifically influence public interest in a species and may not reflect their actual occurrence in a region (Fukano and Soga, 2019; Proulx et al., 2014). For example, *Aesculus hippocastanum* (horse chestnut) is a widely planted street tree gracing many urban parks and thus will be familiar to the general public across the United States despite only being naturalised in a few states. Similarly, crops such as *Ficus carica* (common fig) and *Vitis vinifera* (grape) are grown in most of the United States and will undoubtedly be the subject of Google searches relating to horticultural interests rather than their risk of becoming naturalised. These factors likely contributed to these species being picked up frequently in Google Trends, leading to an overestimate of their actual state-wide occurrence. In contrast, ruderal grasses found in wastelands such as *Aira caryophyllea* (silver hairgrass), *Anthoxanthum odoratum* (sweet vernalgrass), *Bromus sterilis* (brome grass), and *Vulpia bromoides* (brome fescue) were infrequently picked up in Google searches, even though they have naturalised in several states. These short-statured grass species are less noticeable than showy ornamentals or crops and are more difficult to identify, leading to low search interest and an underestimate of their state-wide occurrence in the United States.

Although previous studies have examined the value of Google searches in describing the geographic patterns of alien species in the United States we took the novel step of examining how attributes at a state-level might influence the accuracy with which Google searches could predict the occurrence of alien plant species. We found significant variation across states in terms of the degree to which public searches using Google detected whether a species was present or not in a state. In general, better goodness of fit was found for states with a higher number of alien species and a larger population possessing advanced educational qualifications. A strong correlation between Google search volume and the higher numbers of alien species has been found in previous studies (Fazzari et al., 2024; Fukano and Soga, 2019; Woodworth et al., 2023). With a higher richness of alien species in a region, the public may become more familiar with them through personal encounters or increased media coverage about those species, which can potentially raise search interest. Similarly, people with higher education levels are

likely to have greater knowledge and awareness of alien species and a better understanding of their impacts on ecosystems and biodiversity conservation (Banha et al., 2025; Eiswerth et al., 2011), which in turn can boost public interest in identifying and searching for these species. More educated people are also more likely to have access to online platforms (van den Berg, 2019), which, along with their interest in alien species, can result in more online searches, thus increasing the goodness of fit between the actual occurrence of an alien plant species and that predicted from Google searches. Overall, these findings emphasize the connection between ecological factors and socio-demographic factors in shaping public engagement with biodiversity, including alien plant species in a region, through digital platforms (Novoa et al., 2025).

5. Conclusion

With the increase in alien plant invasions and their associated ecological and economic damages worldwide it is crucial to incorporate innovative and advanced tools and methods for early detection, monitoring, and management of invasions. This study provides insights into how to effectively utilize Google Trends, an underexplored iEcology tool, to enhance understanding of plant invasion ecology. Overall, we demonstrate that for a relatively small subset of alien species, particularly those that are ornamentals or highly invasive, using the scientific name of plants as a keyword can provide some insights into the actual state-wide occurrence of those alien plant species in the United States. Google searches should therefore be interpreted with caution, as they may not always reflect the true distribution of invasive alien plant species. However, they could serve as a potential secondary tool for capturing regional distributions, particularly for species that have high public visibility but are not widespread crops or amenity species.

CRediT authorship contribution statement

Md Azharul Alam: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Philip E. Hulme:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition,

Conceptualization.

Funding

This study was supported by funds provided to Philip E. Hulme from the EU Horizon project OneSTOP (Grant agreement ID: 101180559) and Lincoln University.

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 paper.

Acknowledgements

The authors are grateful to Simon Reynaert, Pavel Pipek, and Ana Novoa for constructive comments on an earlier version of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2026.103689>.

Data availability

The data and code that support the findings of this study are publicly available in Zenodo at the following link: <https://doi.org/10.5281/zenodo.18828189>

References

- Allain, S.J., 2019. Mining Flickr: a method for expanding the known distribution of invasive species. *Herpetol. Bull.* 148, 11–14.
- Altman, A., Shennan, S., Odling-Smee, J., 2022. Ornamental plant domestication by aesthetics-driven human cultural niche construction. *Trends Plant Sci.* 27, 124–138.
- Andrade, C.F.O., Herrera, P.A.L., Bonilla, D.A.G., 2025. Google trends as an early signal in international flower trade. *Ornament. Hort.* 31, e312959.
- Aschim, R.A., Brook, R.K., 2019. Evaluating cost-effective methods for rapid and repeatable national scale detection and mapping of invasive species spread. *Sci. Rep.* 9, 7254.
- Banha, F., Diniz, A.M., Anastácio, P.M., 2025. How people perceive invasion impacts: is there a match between policymakers and stakeholders? *Environ. Sustain. Indic.* 26, 100633.
- Brickell, C., 2011. *American Horticultural Society Encyclopedia of Plants and Flowers*. Dorling Kindersley.
- Caley, P., Cassey, P., 2023. Do we need to mine social media data to detect exotic vertebrate-pest introductions? *Wildl. Res.* 50, 869–875.
- Cebrián, E., Domenech, J., 2023. Is Google trends a quality data source? *Appl. Econ. Lett.* 30, 811–815.
- Cebrián, E., Domenech, J., 2024. Addressing google trends inconsistencies. *Technol. Forecast. Soc. Chang.* 202, 123318.
- Clements, D.R., Upadhyaya, M.K., Joshi, S., Shrestha, A., 2022. *Global Plant Invasions*. Springer Cham, Switzerland.
- Daume, S., 2016. Mining twitter to monitor invasive alien species—an analytical framework and sample information topologies. *Eco. Inform.* 31, 70–82.
- De Hond, A.A., Steyerberg, E.W., Van Calster, B., 2022. Interpreting area under the receiver operating characteristic curve. *Lancet Digit. Health* 4, e853–e855.
- de Mendiburu, F., 2019. *Agricolae: Statistical Procedures for Agricultural Research*. R Package Version 1.3–7.
- Di Lorenzo, P., 2025. *Usmap: US Maps Including Alaska and Hawaii*. R Package Version 0.8.0.
- Dylewski, Ł., Mikula, P., Tryjanowski, P., Morelli, F., Yosef, R., 2017. Social media and scientific research are complementary—YouTube and shrikes as a case study. *Sci. Nat.* 104, 48.
- Eiswerth, M.E., Yen, S.T., van Kooten, G.C., 2011. Factors determining awareness and knowledge of aquatic invasive species. *Ecol. Econ.* 70, 1672–1679.
- ElQadi, M.M., Dorin, A., Dyer, A., Burd, M., Bukovac, Z., Shrestha, M., 2017. Mapping species distributions with social media geo-tagged images: case studies of bees and flowering plants in Australia. *Eco. Inform.* 39, 23–31.
- Fazzari, L., Vardi, R., Jaric, I., Correia, R.A., Coll, M., Sbragaglia, V., 2024. Spatiotemporal patterns of public attention to invasive species across an invasion front: a case study of lionfish (*Pterois miles*) from the Mediterranean Sea. *Biol. Invasions* 26, 3913–3927.
- Fox, J., Weisberg, S., 2018. *An R Companion to Applied Regression*. Sage Publications.
- Foxcroft, L.C., Pyšek, P., Richardson, D.M., Genovesi, P., 2013. Plant invasions in protected areas. Patterns, problems and challenges. *Invading nature—Springer series in invasion. Ecology* 7, 3–18.
- Fukano, Y., Soga, M., 2019. Spatio-temporal dynamics and drivers of public interest in invasive alien species. *Biol. Invasions* 21, 3521–3532.
- González-Moreno, P., Andelković, A.A., Adriaens, T., Botella, C., Demetriou, J., Bastos, R., Bertolino, S., López-Cañizares, C., Essl, F., Fišer, Ž., 2025. Citizen science platforms can effectively support early detection of invasive alien species according to species traits. *People Nat.* 7, 278–294.
- Grömping, U., 2007. Relative importance for linear regression in R: the package relaimpo. *J. Stat. Softw.* 17, 1–27.
- Hartwell, M., Hendrix-Dicken, A.D., Sajjadi, N.B., Bloom, M., Gooch, T., Conway, L., Baxter, M.A., 2022. Trends in public interest in child abuse in the United States: an infodemiology study of Google trends from 2004 to 2022. *Child Abuse Negl.* 134, 105868.
- Hözl, J., Keusch, F., Sajons, C., 2025. The (mis) use of Google trends data in the social sciences—a systematic review, critique, and recommendations. *Soc. Sci. Res.* 126, 103099.
- Jagiello, Z.A., Dyderski, M.K., Dylewski, Ł., 2019. What can we learn about the behaviour of red and grey squirrels from YouTube? *Eco. Inform.* 51, 52–60.
- Jarić, I., Correia, R.A., Brook, B.W., Buettel, J.C., Courchamp, F., Di Minin, E., Firth, J.A., Gaston, K.J., Jepson, P., Kalinkat, G., 2020. iEcology: harnessing large online resources to generate ecological insights. *Trends Ecol. Evol.* 35, 630–639.
- Jarić, I., Bellard, C., Correia, R.A., Courchamp, F., Douda, K., Essl, F., Jeschke, J.M., Kalinkat, G., Kalous, L., Lennox, R.J., 2021. Invasion culturomics and iEcology. *Conserv. Biol.* 35, 447–451.
- Kassambara, A., 2020. *Ggpubr: 'ggplot2' Based Publication Ready Plots*. R Package Version 0.4. 0.
- Kassambara, A., 2021. *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. R Package Version 0.7. 0. [Computer Software. https://CRAN.R-project.org/package=rstatix](https://CRAN.R-project.org/package=rstatix).
- Leighton, G.R., Hugo, P.S., Roulin, A., Amar, A., 2016. Just Google it: assessing the use of Google images to describe geographical variation in visible traits of organisms. *Methods Ecol. Evol.* 7, 1060–1070.
- Liu, K., 2024. The measurement errors of google trends data. *Discov. Data* 2 (1), 7.
- Lolić, I., Matošec, M., Sorić, P., 2024. DIY google trends indicators in social sciences: a methodological note. *Technol. Soc.* 77, 102477.
- Maciejewski, C., Lippitt, K., Ringholm, E., Katzenmoyer, J., Van Huynh, A., 2025. Using Google trends to improve monitoring of the invasive spotted lanternfly in the United States. *Biol. Invasions* 27, 150.
- Mackey, T., Baur, C., Eysenbach, G., 2022. Advancing infodemiology in a digital intensive era. *JMIR Infodemiol.* 2 (1), e37115.
- Marcenò, C., Padullés Cubino, J., Chytrý, M., Genduso, E., Salemi, D., La Rosa, A., Gristina, A.S., Agrillo, E., Bonari, G., Giusso del Galdo, G., 2021. Facebook groups as citizen science tools for plant species monitoring. *J. Appl. Ecol.* 58, 2018–2028.
- Mavragani, A., 2020. Infodemiology and infoveillance: scoping review. *J. Med. Internet Res.* 22 (4), e16206.
- Mavragani, A., Ochoa, G., 2019. Google trends in infodemiology and infoveillance: methodology framework. *JMIR Public Health Surveill.* 5, e13439.
- Meyer, D., Zeileis, A., Hornik, K., 2020. *vcd: Visualizing Categorical Data*. R Package Version 1.4-8. Recuperado de <https://rdr.io/cran/vcd>.
- Miranda, E.B., Ribeiro Jr., R.P., Strüssmann, C., 2016. The ecology of human-anaconda conflict: a study using internet videos. *Trop. Conserv. Sci.* 9, 43–77.
- Mittermeier, J.C., Correia, R., Grenyer, R., Toivonen, T., Roll, U., 2021. Using Wikipedia to measure public interest in biodiversity and conservation. *Conserv. Biol.* 35, 412–423.
- Nininahazwe, F., Théau, J., Marc Antoine, G., Varin, M., 2023. Mapping invasive alien plant species with very high spatial resolution and multi-date satellite imagery using object-based and machine learning techniques: a comparative study. *GISci. Remote Sens.* 60, 2190203.
- Novoa, A., Jarić, I., Pipek, P., Pyšek, P., 2025. Culturomics and iEcology provide novel opportunities to study human and social dimensions of alien species introductions. *Trends Ecol. Evol.* 40, 18–26.
- Pedersen, T.L., 2019. Patchwork: The Composer of Plots. CRAN: Contributed Packages.
- Pervaiz, F., Pervaiz, M., Rehman, N.A., Saif, U., 2012. FluBreaks: early epidemic detection from Google flu trends. *J. Med. Internet Res.* 14 (5), e2102.
- Pocock, M.J., Adriaens, T., Bertolino, S., Eschen, R., Essl, F., Hulme, P.E., Jeschke, J.M., Roy, H.E., Teixeira, H., De Groot, M., 2024. Citizen science is a vital partnership for invasive alien species management and research. *IScience* 27, 108623.
- Polak, Y.E., 2025. Two heads are better than one. *Autom. Document. Math. Linguist.* 59, S37–S46.
- Proulx, R., Massicotte, P., Pepino, M., 2014. Googling trends in conservation biology. *Conserv. Biol.* 28, 44–51.
- R Core Team, 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Rakgoale, P.B., Ngetar, S.N., 2024. Detecting invasive alien plant species using remote sensing, machine learning and deep learning: a systematic review. *J. Sens.* 2024, 8854675.
- Revelle, W., 2019. *Procedures for Psychological, Psychometric, and Personality Research* [R Package Psych Version 1.8.12]. R-Project. Org.
- Ripley, B., 2011. *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. R Package Version 7, pp. 3–29.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., Müller, M., 2011. *PROC: an open-source package for R and S+ to analyze and compare ROC curves*. *BMC Bioinformatics* 12, 77.

- Rovetta, A., 2024. Google trends in infodemiology: methodological steps to avoid irreproducible results and invalid conclusions. *Int. J. Med. Inform.* 190, 105563.
- Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., Toomet, O., Crowley, J., Hofmann, H., 2021. GGally: Extension to ggplot2.
- Seifter, A., Schwarzwald, A., Geis, K., Aucott, J., 2010. The utility of "Google trends" for epidemiological research: lyme disease as an example. *Geospat. Health* 4 (2), 135.
- Springer, S., Strzelecki, A., Holmes, B., Ziermann-Canabarro, J.M., Kaatz, M., Zieger, M., 2025. Setting trends with google: limits and perspectives when utilising search engine data. *Qual. Quant.* <https://doi.org/10.1007/s11135-025-02431-0>.
- van den Berg, A.C., 2019. Participation in online platforms: examining variations in intention to participate across citizens from diverse sociodemographic groups. *Perspect. Public Manage. Govern.* 4, 259–276.
- Vardi, R., Mittermeier, J.C., Roll, U., 2021. Combining culturomic sources to uncover trends in popularity and seasonal interest in plants. *Conserv. Biol.* 35, 460–471.
- Vilà, M., Hulme, P.E., 2017. Impact of Biological Invasions on Ecosystem Services. Springer.
- Wang, A., McCarron, R., Azzam, D., Stehli, A., Xiong, G., DeMartini, J., 2022. Utilizing big data from google trends to map population depression in the United States: exploratory infodemiology study. *JMIR Mental Health* 9, e35253.
- Wei, T., Simko, V., 2024. R Package "Corrplot": Visualization of a Correlation Matrix. 487 (Version 0.94).
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D.A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., 2019. Welcome to the Tidyverse. *J. Open Source Softw.* 4, 1686.
- Wilson, J.R., Panetta, F.D., Lindgren, C., 2017. Detecting and Responding to Alien Plant Incursions. Cambridge University Press.
- Wong, M.K., 2024. Misrepresentation of invasive species in the mass media with images of unrelated organisms. *Conserv. Biol.* 38, e14382.
- Woodworth, E., Tian, A., Blair, K., Pullen, J., Lefcheck, J.S., Parker, J.D., 2023. Media myopia distorts public interest in US invasive plants. *Biol. Invasions* 25, 3193–3205.
- Yang, C.-X., Baker, L.M., McLeod-Morin, A., 2024. Trending ticks: using Google trends data to understand tickborne disease prevention. *Front. Public Health* 12, 1410713.