



INFERRING THE INVASIVE STAGE OF EURASIAN SPIDER *MYRMARACHNE FORMICARIA* (ARANAEA: SALITICIDAE) IN NORTH AMERICA USING SPECIES DISTRIBUTION MODELS

JOSEPH GIULIAN^{1,*}

¹ Department of Biological Sciences, East Tennessee State University, Johnson City, TN, USA

* Corresponding author: Joseph Giulian; E-mail address: giulian@etsu.edu;
Phone: [+1-5185679818](tel:+15185679818)

Abstract.

Introduced to North America sometime before 2001, *Myrmarachne formicaria*'s potential for expansion and its current invasion stage are yet to be estimated. MaxEnt species distribution models (SDMs) for this species were developed with open-source data to predict suitable regions in North America and globally. Comparisons of SDM predictions across global, native, and invasive occurrences were made to hypothesize the invasion stage for North American populations. Suitable habitats were identified from the east coast of North America, extending west to Indiana and Michigan, and south to southern Appalachia. Additionally, regions of South America, Africa, Australia, and New Zealand, fall within its fundamental climate niche. Predictions show invasive *M. formicaria* populations occupy stabilized and colonizing habitats, suggesting an ongoing colonizing phase with high risk for further spread. All SDMs showed a high sensitivity to temperature variables, with the invasive SDM uniquely sensitive to precipitation. The prevalence of stabilized invasive populations implies niche conservatism and suggests that niche differences more likely reflect specialization rather than a niche shift. Overall, findings indicate that *M. formicaria* has extensive climatically suitable habitat both within and outside of North America. Further research and the development of targeted mitigation strategies are imperative to prevent its spread to these environments.

Keywords: Biological invasion, invasive, spider, climate niche, invasion stage

INTRODUCTION

The Eurasian salticid spider, *Myrmarachne formicaria* (de Geer 1778), is a recent adventive species in North America. Since its 2001 detection in Ohio, USA, the spider has been collected from Michigan, Pennsylvania, New York, and Ontario, Canada (Bradley et al. 2006; Gall and Edwards 2016; GBIF 2024, *M. formicaria*). Historically, its known range encompasses Europe, Russia, Caucasus, Turkey, Iran, China, Korea, and Japan (Kim and Lee 2014; World Spider Catalog 2023). Specifics surrounding the North American introduction remain unclear, but it is speculated that *M. formicaria* was inadvertently imported with horticultural products (Bradley et al. 2006). The spider displays a preference for sunny, low vegetation areas with ground debris, but appears somewhat synanthropic in its newfound range (Bradley et al. 2006; Buchanan and Bradley 2012; Gall and Edwards 2016). Its ant mimicry, involving locomotor behavior (Shamble et al. 2017) and possibly hydrocarbon mimicry (Pekár and Jiroš 2011), also reduces predation risks, likely enhancing this spider's ability to invade and exploit new environments. Documenting the spread of invasive spiders is crucial due to their ability to alter local ecology through predation on arthropods and competition with native species (Nentwig

2015). Of particular interest is this spider's potential to compete for the same functional (e.g. "Eltonian") niche as native myrmecomorphs. As Batesian mimics (Shamble et al. 2017), myrmecomorph spiders benefit from predators' learned avoidance of their ant models, an effect that could be diminished by the introduction of a new mimic (Lindström et al. 1997). This stands to reduce the mimicry system's effectiveness, thereby increasing predation risks for native myrmecomorphs. Despite these potential effects from a spider apparently spreading throughout its newfound range (Bradley et al. 2006; Gall and Edwards 2016; GBIF 2024, *M. formicaria*), the ecological impacts and the full extent of its spread remain unassessed.

Species distribution models (SDMs) remain one of the most useful tools for forecasting species distributions (Guisan and Thuiller 2005; Barbosa et al. 2012). SDMs can be trained with environmental and species occurrence data to help identify suitable habitats, predict potential ranges, and quantify environmental niches (Vetaas 2002; Guisan and Thuiller 2005; Elith and Leathwick 2009; Barbosa et al. 2012; Giulian et al. 2024). Open-source environmental and occurrence data have become highly accessible in recent years, providing ample resources to probe the abiotic drivers of species distributions. Data derived

from precipitation and temperature data are commonly used for SDM research, especially a group of 19 biologically relevant climatic variables developed for ecological applications (O'Donnel and Ignizio 2012). Citizen science and museum databases also offer a highly accessible route to occurrence data (Ivanova and Shashkov 2021). Unfortunately, occurrence data are often presence-only data. This presents an interpretational challenge for SDM methods that assume knowledge of areas where the species is absent. However, open-source presence-only data can still be used to produce valuable SDM assessments (Elith et al. 2006). 'MaxEnt' is among the most common choices for modeling presence-only data (Phillips et al. 2006; Elith et al. 2010; Phillips et al. 2017). It uses a 'maximum entropy' approach to estimate the probability distribution of a species occurrence across space (Phillips et al. 2006; Elith et al. 2010; Phillips et al. 2017). These SDMs are also highly robust to collinearity issues due to inbuilt regularization methods (Phillips et al. 2006; Hernandez et al. 2006; Elith et al. 2010; Feng et al. 2019). Altogether, MaxEnt's capabilities make it an effective tool for building SDMs using open-source data.

A fundamental assumption of SDMs is that the species being modeled inhabits all suitable areas and is absent from all unsuitable ones (Guisan and Thuiller 2005). This assumption is especially violated in the case of invasive species ranges. However, SDMs trained using invasive occurrence data can still allow for hypotheses and inference into the stages of invasion (Gallien et al. 2012; Barbosa et al. 2012; Cabra-Rivas et al. 2016; Taucare-Ríos et al. 2016; Taucare-Ríos et al. 2018; Giuliani et al. 2024). SDMs trained to specific regions of occurrence reflect the realized niches of those regions, while SDMs trained using the entire range of occurrence are a proxy for the fundamental niche (Vetaas 2002; Gallien et al. 2012; Taucare-Ríos et al. 2018). When regions are predicted as suitable by both global and regional SDMs, it suggests these areas support stabilizing populations (Gallien et al. 2012; Taucare-Ríos et al. 2018; Giuliani et al. 2024). Populations in stabilizing habitat are the most likely to have practiced niche conservation during invasion (Taucare-Ríos et al. 2018). Alternatively, regions identified as suitable by a global SDM but not by a regional SDM are expected to be undergoing a colonizing phase, where the species is expanding within the fundamental niche, but are not yet fully adapted or stabilized (Gallien et al. 2012; Taucare-Ríos et al. 2018). Areas predicted as suitable by a regional SDM but not by the global SDM would instead suggest the species requires local adaptation in those habitats (Gallien et al. 2012; Taucare-Ríos et al. 2018). Any occurrences outside of predictions are therefore sink populations. Using this conceptual framework, SDMs can be trained with invasive, native, and all occurrences to create more holistic and nuanced hypotheses about the invasive process. These methods have been previously applied to infer the potential ranges and stages of invasion for seven plant species and three spider species (Gallien et al. 2012; Cabra-Rivas et al. 2016; Taucare-Ríos et al. 2016; Taucare-Ríos et al. 2018; Giuliani et al. 2024).

SDMs for *M. formicaria* were developed using open-source environmental and occurrence data from native, invasive, and combined populations. Predictions allowed for a detailed evaluation of areas most likely to support its climate niche, especially in North America. Through comparing the predictions from all three SDM outputs, regions could also be identified that support colonizing, adapting, or stabilizing phases of invasion, shedding light on the potential extent of spread and range-filling for *M. formicaria*.

METHODS

Occurrence data

Open-source presence-only data for *M. formicaria* was sourced from the Global Biodiversity Information Facility (GBIF) as of 25 December 2023 (GBIF Secretariat 2022; GBIF 2024, *M. formicaria*). The timeframe for occurrences was 1 January 1988, to 25 December 2023. To enhance data quality, 'CoordinateCleaner' was used to remove problematic data (Zizka et al. 2019). This R package eliminated duplicates, imprecise occurrence data, and occurrences located at centroids of countries or capitals. It also removed data with zero coordinates, appearing in the ocean, or representing geographic outliers (Zizka et al. 2019). To address spatial biases, the R package 'spThin' (Aiello-Lammens et al. 2015) was applied for spatial thinning of occurrences to 20 km. 143 invasive and 235 native occurrences were deemed acceptable after pre-processing procedures.

Predictors

Open-source monthly temperature and precipitation data were taken at 30 arcsecond resolution from CHELSA (Karger et al. 2018). These data spanned 1 January 1988 to 31 December 2018 (Giulian et al. 2024). The R package 'dismo' was used to derive 19 bioclimatic variables from the monthly climate data (O'Donnel and Ignizio 2012; Hijmans et al. 2017; Giuliani et al. 2024). All variables were then scaled to 2.5 arcminute resolution (~4 km at the equator). This choice reduced: a) computational requirements, b) the chance that an occurrence was misplaced in an adjacent cell, and c) the risk of over-fitting, which each come with using finer resolutions. Given the wide-spanning range of *M. formicaria*, the resolution of 2.5 arcminutes is expected to still reflect environmental patterns that affect its distribution.

Variable selection followed an impartial procedure where all variables were considered potentially biologically relevant. A statistical approach was then used to ascertain the combination of variables with minimum collinearity. Using this method allowed for the objective selection of environmental variables, while also reducing variable redundancy by better ensuring each abiotic variable independently contributed to describing environmental variation. Variable inflation factors (VIFs) and Pearson correlation coefficients were the selection criteria. The 'usdm' R package identified the largest set of variables where each

had a VIF lower than 5 and a Pearson coefficient below 0.70 (Naimi et al. 2014).

The variables deemed acceptable include annual mean temperature (bio1), isothermality (bio3), maximum temperature of the warmest month (bio5), mean temperature of the wettest quarter (bio8), precipitation seasonality (CV) (bio15), precipitation of the warmest quarter (bio18), and precipitation of the coldest quarter (bio19). Annual mean temperature, maximum temperature of the warmest month, and mean temperature of the wettest quarter are measured in degrees Celsius. Precipitation of the warmest and coldest quarters are in millimeters. Isothermality is the ratio of day-to-night temperature fluctuation to summer-to-winter fluctuation, where a value of 100 indicates equal diurnal and annual ranges, and values below 100 indicate less monthly than yearly variability (O'Donnel and Ignizio 2012). Precipitation seasonality is also a percentage as it is the coefficient of variation of annual precipitation.

Model workflow

Using MaxEnt v3.4.4 (Phillips et al. 2006; Phillips et al. 2017) and the R package 'ENMeval' v2.0.4 (Muscarella et al. 2014), SDMs were developed across three independent workflows, each trained with either invasive, native, or all (global) occurrences. Four-fold spatial block cross-validation was used for SDM training (Muscarella et al. 2014). Training regions were confined to a 500 km radius around occurrences. The dataset was divided into four geographically distinct blocks for each workflow, with SDMs iteratively trained on three blocks and tested on the fourth. Consequently, four SDMs were built for each SDM type, and the average of these four was taken to represent the final SDM. SDM types of various combinations of MaxEnt's adjustable regularization multipliers and feature classes were considered. Response features chosen were linear, quadratic, hinge, linear-quadratic, and linear-quadratic-hinge, with regularization multipliers set from 1 to 5 in 1-unit increments. This led to 25 unique SDM types for each of the three regional extents. The SDM type with the lowest information criterion (AICc) was then selected for each extent. AICc is a useful metric for selecting more accurate MaxEnt SDMs, especially for smaller datasets (Warren and Seifert 2011). Lower AICc values suggest SDMs that more effectively capture data trends without using excessive parameters, thus minimizing the risk of overfitting and avoiding unnecessary complexity.

To examine whether SDMs selected by AICc performed significantly better than random, selected SDMs were compared to null SDMs built using the same cross-validation methods and spatial extents (Bohl et al. 2019; Kass et al. 2020). Training data for null SDMs was randomly sampled from each spatial block and equals the number of *M. formicaria* occurrences for that fold (Kass et al. 2020). Each null SDM was then tested using the real training data of the withheld fold (Bohl et al. 2019; Kass et al. 2020). This was repeated 1000 times for each selected SDM to build null distributions for area under the curve (AUC) values. Native, invasive and global SDM AUCs

were then compared against null AUC distributions. A one-tailed T test determined whether selected SDMs had significantly better discrimination ability than random (Kass et al. 2020). All null SDMs and related comparisons were made using the R package ENMeval v2.0.4 (Muscarella et al. 2014).

SDM predictions for three regional extents are provided. Predictions reflect relative occurrence rates (ROR), a proxy for habitat suitability that ranges from 0 to 1 (Phillips et al. 2006; Phillips et al. 2017). SDMs were trained on and transferred to native, invasive and global regions to visualize areas of habitat suitability. Clamping was deactivated to prevent any SDM from extrapolating predictions to environments outside the range of its training data. To facilitate comparison, predictions were converted to binary maps using an ROR of 0.50 for minimum suitability (Gallien et al. 2012; Taucare-Ríos et al. 2016; Giulian et al. 2024). Habitats predicted only by the invasive and native SDMs are where populations would require local adaptation, while those predicted solely by the global SDM are inferred to support colonizing populations. Regions predicted as suitable by both global and invasive or native SDMs support the most stabilized populations; all remaining regions are classified as sink habitats. Results are provided in niche and geographical space.

Variable permutation importance was reported and compared across SDMs. For Maxent SDMs, permutation importance is determined by randomly permuting each environmental variable using the training data, then assessing the impact on SDM performance (Phillips et al. 2006; Kass et al. 2020). This method focuses solely on the final SDM, independent of its development path (Phillips et al. 2006), to present the most important predictors of species distribution.

Model caveats

Though MaxEnt SDMs offer a sophisticated method for handling presence-only data, they have limitations. Data inaccuracies can lead to erroneous MaxEnt predictions. This risk was mitigated by selecting accurate occurrence data and reliable environmental data. However, presence-only datasets still often have spatial biases due to uneven sampling and reporting efforts. This, combined with uncertainty about 'absence' sites, can cause MaxEnt SDMs to underrepresent undersampled regions and overfit oversampled ones. To reduce spatial clustering effects, occurrences were thinned to 20 km. Despite these efforts, undersampled regions may still be underrepresented in final predictions. Furthermore, biotic interactions are not considered in these SDMs, so predictions do not account for competition, symbioses, anthropogenic effects, etc., which could further limit or support invasive success. At the resolution of 2.5 arcminutes, they may also be unlikely to capture fine-scale trends, instead capturing large-scale trends in the species distribution. As such, efforts were made to draw conclusions from major patterns and similarities across SDM.

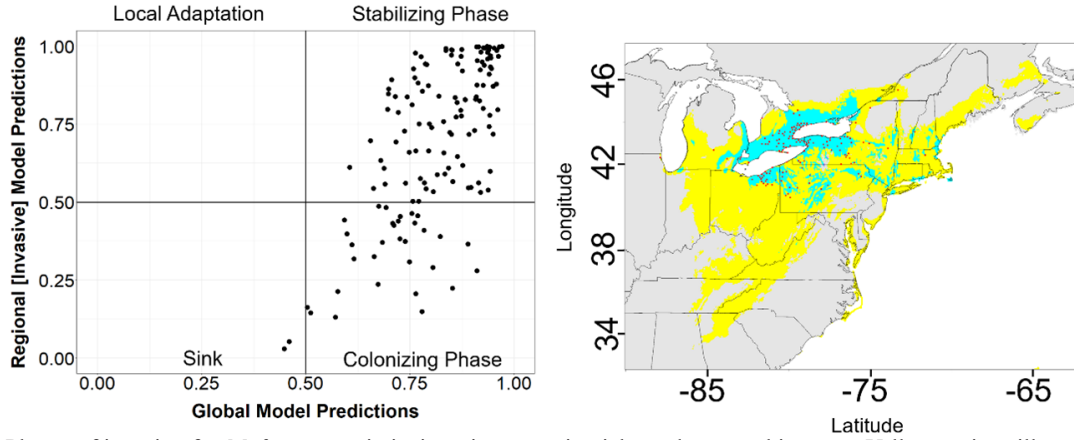


Figure 1. Phases of invasion for *M. formicaria* in its invasive range in niche and geographic space. Yellow regions illustrate habitat supporting the colonizing phase (e.g. regions predicted only by the global SDM, approximating the fundamental niche). Cyan regions illustrate the stabilizing phase (e.g. regions of agreement between the global and regional models, approximating the overlap of fundamental and realized niches). Red points reflect invasive occurrences.

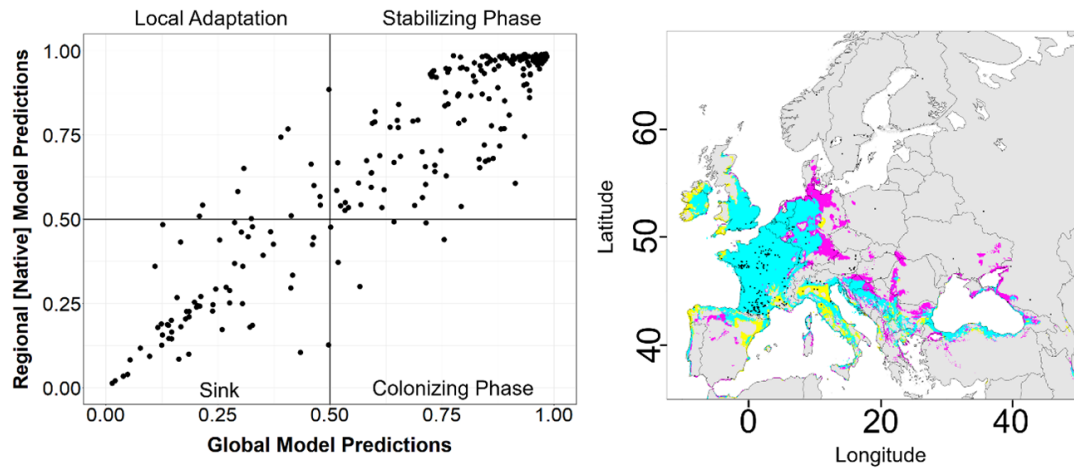


Figure 2. Phases of invasion for *M. formicaria* in its native range in niche and geographic space. Yellow regions reflect habitat supporting the colonizing phase, magenta the local adaptation phase (e.g. regions predicted only by regional models, approximating the realized niches), and cyan the stabilizing phase (both SDMs). Black points reflect native occurrences.

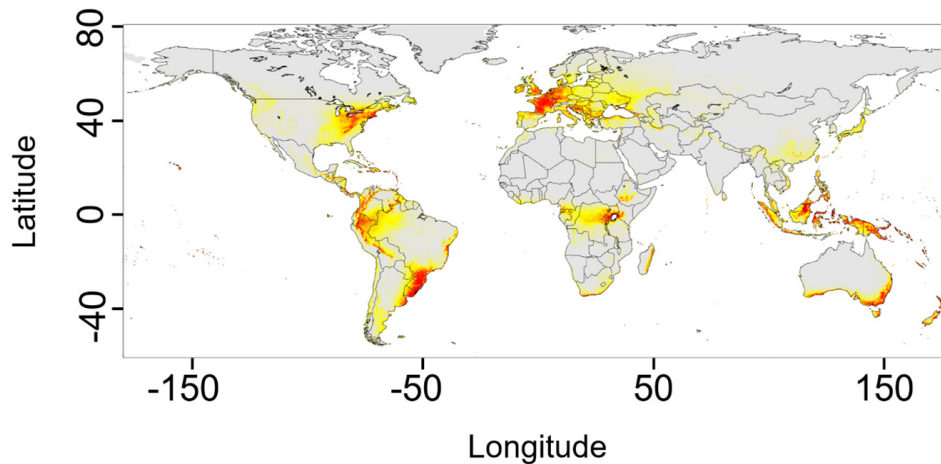


Figure 3. World map of climatically suitable habitat predicted by the global SDM. For the suitability (ROR) scale from 0 to 1, 0 is grey, the lowest nonzero values begin at yellow, intermediate suitability is orange, and maximum suitability is red.

Table 1. Permutation importance of variables used to SDM global, invasive and native *M. formicaria* occurrences.

| Permutation Importance (percent) | | | |
|----------------------------------|--------|----------|--------|
| Variable | Global | Invasive | Native |
| Annual Mean Temp. | 52.96 | 30.80 | 51.00 |
| Isothermality (%) | 2.05 | 4.83 | 4.32 |
| Max Temp. Warmest Month | 20.16 | 11.15 | 25.15 |
| Mean Temp. Wettest Quarter | 0.90 | 0.26 | 7.69 |
| Precip. Seasonality (CV) | 12.38 | 36.90 | 4.51 |
| Precip. of Wettest Quarter | 0.00 | 1.48 | 5.66 |
| Precip. of Coldest Quarter | 11.55 | 14.58 | 1.67 |

RESULTS

The distribution of *M. formicaria* is primarily shaped by temperature. Annual mean temperature was the main driver for native and global SDMs and the second most important for the invasive SDM (Table 1). Suitability increased with annual mean temperature for both global and native SDMs but peaked at 11.1°C for the invasive SDM. The maximum temperature of the warmest month also had a strong effect on all SDMs, indicating that higher maximum temperatures indeed impose constraints on the species' distribution. For the global SDM, suitability increased with maximum temperature of the warmest month until 26.0°C, after which it decreased. For the native model, suitability decreased beyond 24.9°C, and for the invasive model, it decreased beyond 22.1°C.

Invasive populations show heightened sensitivity to precipitation. Precipitation seasonality predominantly influenced the invasive SDM but was less important for global and native SDM (Table 1). Still, all SDMs showed decreasing suitability with increasing precipitation seasonality. This reveals that *M. formicaria* finds greater success in habitats with less seasonal variation in precipitation, especially during its invasion. Thus, habitats with increasing seasonal variation are predicted to pose a greater challenge for the establishment and spread of this species, easing the control measures necessary for those regions. Other variables such as isothermality, mean temperature of the wettest quarter, and precipitation of the warmest quarter, had minimal effects across all SDMs. The AICc values for the native, invasive, and global SDMs were 5805, 2878, and 9598, respectively (Supplemental Table 1). All three SDMs had significantly better discrimination ability than null SDMs (Supplemental Table 1; Supplemental Figs. 1, 2, and 3).

In North America, *M. formicaria* appears stabilized and actively colonizing climatically permissive habitat. The majority (77%) of invasive occurrences were in stabi-

lizing habitat while the rest occurred in colonizing (22%) or sink (1%) habitat (Fig. 1). None were in regions requiring local adaptation. Stabilizing habitat centralized along the 42nd parallel, mainly within the Lake Erie Lowland, Erie Drift Plain, and Eastern Great Lakes and Hudson Lowlands ecoregions. These ecoregions, primarily composed of temperate broadleaf forests, mixed forests, and agricultural lands, reflect climatologies within both the realized and fundamental niches. Invasive populations here should be the hardest to destabilize and are likely sources of future founder events, especially to surrounding habitats still within the fundamental climatic niche. Colonizing habitats were predicted for the surrounding Eastern Temperate Forest and Northern Forest ecoregions across New England, the Mid-Atlantic, and some Midwestern and Southern states. Regions outside of these predictions are inferred to be least climatically permissive, more likely serving as climatic barriers to the spread of *M. formicaria*.

As found for North American predictions, the native predicted range for *M. formicaria* primarily spans temperature deciduous and mixed forest ecoregions. Stabilizing native habitat was predicted in Southern Temperate Atlantic Mixed Forest, Northern Temperate Atlantic Mixed Forest, and Western European Broadleaf Forest ecoregions (Fig. 2). Countries predicted to host stabilizing native habitat include France, Spain, United Kingdom, Belgium, Netherlands, Luxembourg, Germany, Italy, Slovenia, Croatia, Serbia, and Bosnia and Herzegovina. The majority (67%) of occurrences were in stabilizing habitat, but 25% occurred in sink habitat, 6% in habitat requiring local adaptation, and 3% in colonizing habitat (Fig. 2). Local adaptation occurred in parts of Germany, the Balkans, and along the Iberian Peninsula. Except for the Balkans, *M. formicaria*'s native occurrences and climate niche predictions appear to be associated with lower elevations.

Regions of suitability within the fundamental climate niche are also forecasted outside of the current

range. Global predictions show areas of tropical and subtropical broadleaf wet forests offer climatically permissive habitat for *M. formicaria* in South America, Oceania, eastern Madagascar, and parts of central Africa (Fig. 3). These ecoregions are distinctly different floristically and seasonally than those encompassing current native and invasive climate niches. Temperate broadleaf and mixed forests of New Zealand and southern Australia, which bear greater similarity to the ecoregions occupied by known populations, were also forecasted to offer the fundamental climate niche.

DISCUSSION

Overall, there is a high risk for continued spread of *M. formicaria* in North America and globally. Temperate forest ecoregions host the majority of *M. formicaria* occurrences and predictions, though its fundamental climate niche appears in some subtropical forest regions. Invasive *M. formicaria* populations are either stable or colonizing, implying the spider is actively colonizing new habitats rather than having occupied its entire equilibrium range or constituting locally adapted or sink populations (Gallien et al. 2012). Habitats predicted to support this colonizing phase extend to the east coast of North America, west to Indiana and Michigan, and south to southern Appalachia. Assuming factors beyond climate do not bar the species from succeeding in suitable yet uninhabited areas, there is high potential for continued colonization and range filling.

Integrating climatic predictions into control efforts allows for more realistic and targeted strategies. Populations in stabilizing habitats experience limited climate constraints, making these areas the primary focus for control measures to prevent the spread from source populations. In contrast, efforts to destabilize surrounding areas supporting other invasive phases should require less substantial interventions. Habitats that support the colonizing phase still fall within the fundamental climate niche, but various biotic or abiotic factors could be preventing this niche from becoming fully realized. These regions are comparatively less likely to immediately support large, stable source populations, making control efforts in colonizing habitats more effective and less resource-intensive than in stabilizing habitats. Sink populations outside of climatically permissive areas are the least likely to sustain the species long-term, making these regions the least likely to require intensive management plans.

The unique biology of *M. formicaria* makes its potential invasive ecological interactions multifarious. As an arthropod predator, it can possibly drive fundamental shifts in species compositions through predation on other arthropods, as well as through intraguild predation and competition with other arthropod predators (Nentwig 2015). The myrmecomorphic morphology and behavior of this species (Pekár and Jiroš 2011; Shamble et al. 2017) specifically enhance its invasive potential. Bearing resemblance to an ant is an adaptive Batesian mimicry strategy that reduces predation in *M. formicaria* (Shamble et al. 2017), supporting survival and aiding its spread across hospitable

habitats. Another Eurasian ant-mimic spider, *Synageles venator* (Lucas), has already found success invading regions of eastern southern Canada (Bradley et al. 2006). The introduction of new mimics can profoundly decrease mimic survival through predator confusion (Lindström et al. 1997). Numerous native myrmecomorphic salticid spiders currently exist in eastern North America, such as those in the genera *Peckhamia* Simon, *Sarinda* Peckham & Peckham, and *Synemosyna* Hentz. These native myrmecomorphs appear to be major competitors for the Eltonian niches being invaded *M. formicaria* and *S. venator*, and the most likely candidates to experience decreased benefits from their Batesian mimicry strategies now and in the future.

It is also possible that myrmecomorphy in *M. formicaria* supports the aggressive mimicry of ants. However, it is unclear to what degree *M. formicaria* is myrmecophilous and whether individuals play a role as inquilines or predators of ants. The spider indeed bears a striking resemblance to *Formica rufa* L. and is known to associate with *Formica* colonies in its invasive range (Bradley et al. 2006). In fact, over 50 species of native and invasive *Formica*, including the Eurasian *F. rufa*, occur within the predicted invasive range for *M. formicaria* (GBIF 2024, *Formica*). If symbioses with *Formica* support its success, invasive *M. formicaria* will have plentiful opportunity to associate with ants of this genus to support continued spread.

Native populations of *M. formicaria*, which should reflect a range nearer to equilibrium, were largely in stable or in sink habitats. This is a range-filling pattern observed in native populations of the invasive spider *Loxosceles rufescens* (Dufour) (Taucare-Ríos et al. 2018). For the synanthropic *L. rufescens*, its niche-space pattern is probably best explained by frequent dispersal from suitable habitats through human activities (Taucare-Ríos et al. 2018). The synanthropic tendencies of *M. formicaria* (Bradley et al. 2006; Gall and Edwards 2016) should similarly support its colonization of new habitats via human-mediated dispersal. Most native *M. formicaria* sink populations are situated far from the stabilized range in populated areas along the Scandinavian coastline, probably having dispersed there by anthropogenic means. Human-mediated dispersal facilitated *M. formicaria*'s introduction to North America and is likely to promote its continued expansion into new environments.

Permutation variable importance revealed that native and invasive niches are shaped by slightly different conditions. Fundamental and realized climate niches are all highly sensitive to temperature; however, the invasive realized niche appears more adapted to hydrological conditions than the native realized niche. New environments sometimes impose enough abiotic or biotic pressure to compel a niche shift (Guisan et al. 2014; Taucare-Ríos et al. 2016; Taucare-Ríos et al. 2018). However, this scenario appears unlikely. Fundamental niche constraints are difficult to overcome, and only 10% of the world's top invasive species have shown evidence of such a niche shift

(N.A. et al. 2022). Individuals in a stabilizing habitat are likely to have practiced niche conservation during invasion, given these regions represent the theoretical overlap space of fundamental and realized niches (Taucare-Ríos et al. 2018). Since most invasive occurrences are indeed observed in stabilizing habitat, the invasive realized niche for *M. formicaria* is likely conserved within the bounds of its fundamental niche. Differences in variable importance are therefore better attributed to regional niche specialization rather than a niche shift.

Other regions outside the current range of occurrence are predicted to be at-risk for invasion. Global SDM predictions show the fundamental climate niche for *M. formicaria* occurs in South America, Oceania, New Zealand, and Australia. Spiders and other arthropods of European or Eurasian origin have already been successfully introduced to New Zealand (Vink and Duperre 2011; Vink and Kean 2013), southern Australia (Belosludtsev and Gasilin 2018; Quarrell et al. 2018), and South America (Wyckhuys et al. 2013; Dupérré 2023). The high potential for continued range-filling within and outside of North America for *M. formicaria* should prompt focus on determining its causal mechanisms of dispersal. Such knowledge could inform mitigation strategies to reduce or inhibit the spider's spread to other suitable habitat. Attention may best be directed towards its potential transport via horticultural products, the speculated mechanism of entry to North America (Bradley et al. 2006). Although it is yet unclear what aids its dispersal or what effect its spread will have on the ecology of North American habitats, SDM forecasts show plentiful habitat for colonization by *M. formicaria* in North America and beyond.

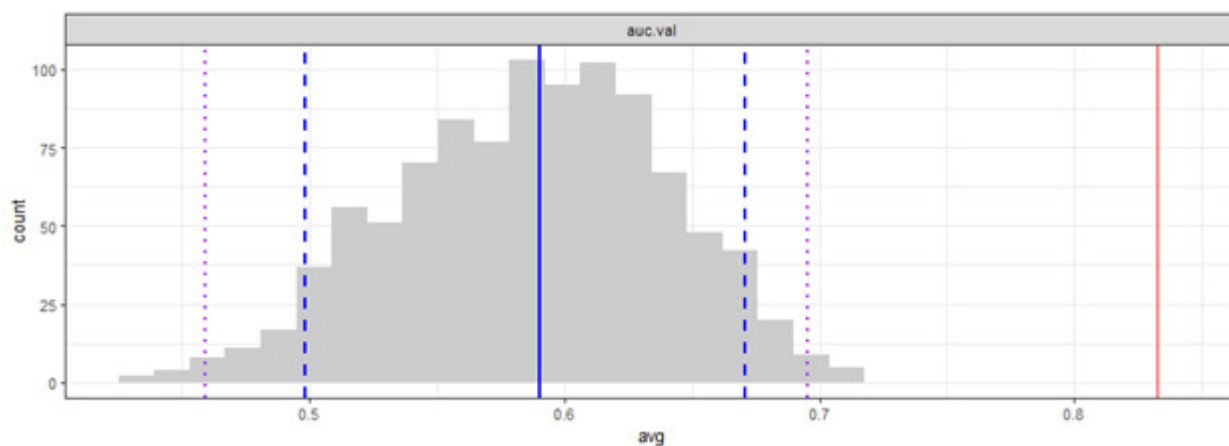
ACKNOWLEDGEMENTS

I am grateful to the faculty of the Biomedical Sciences and Biological Sciences departments for their continued support. Specifically, thank you to Drs. Thomas C. Jones and Darrell Moore, who offered thoughtful discussion and edits that enhanced the manuscript's quality.

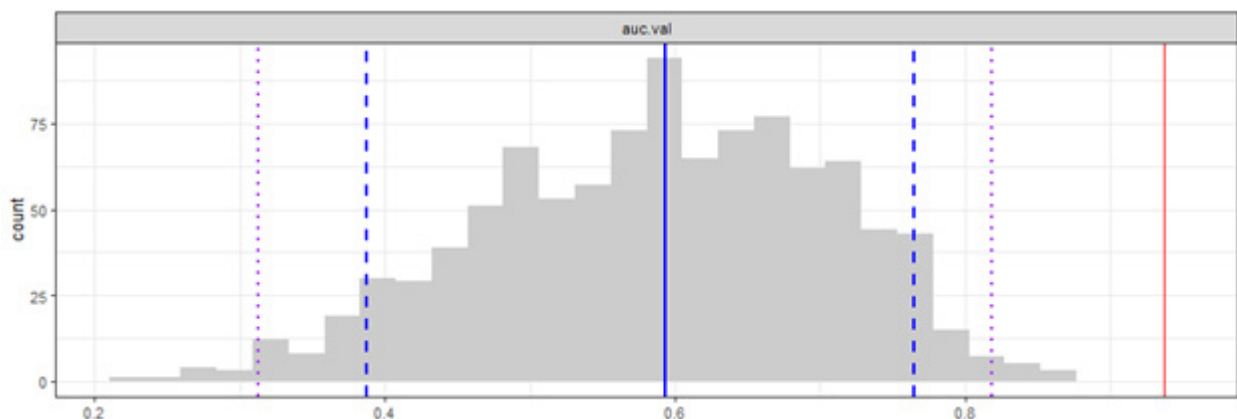
REFERENCES

- Barbosa F.G., Schneck F. & Melo A.S. (2012) Use of ecological niche models to predict the distribution of invasive species: a scientometric analysis. *Brazilian Journal of Biology*, 72, 821-829 <https://doi.org/10.1590/S1519-69842012000500007>
- Belosludtsev, E. A. & Gasilin, V. V. (2018) Cosmopolitan distribution of the spitting spider *Scytodes thoracica* Latreille, 1802 (Aranei, Scytodidae) and its new findings in the Northern Parts of the Range. *Contemporary Problems of Ecology*, 11, 123-136. <https://doi.org/10.1134/S1995425518020038>
- Bradley, R. A., Cutler, B. & Hodge, M. (2006) The first records of *Myrmarachne formicaria* (Araneae, Salticidae) in the Americas. *Journal of Arachnology*, 34, 483-484. <https://doi.org/10.1636/H04-67.1>
- Buchanan, S. & Bradley, R. A. (2012) *Common Spiders of North America*. University of California Press.
- Cabra-Rivas, I., Saldaña, A., Castro-Díez, P. & Gallien, L. (2016) A multi-scale approach to identify invasion drivers and invaders' future dynamics. *Biological Invasions*, 18, 411-426. <http://dx.doi.org/10.1007/s10530-015-1015-z>
- Dupérré, N. (2023) *Araneae (spiders) of South America: a synopsis of current knowledge*. *New Zealand Journal of Zoology*, 50(1), 3-17. <https://doi.org/10.1080/03014223.2021.2022722>
- Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129-151. <https://doi.org/10.1111/j.2006.0906-7590.04596.x>
- Elith, J. & Leathwick, J. R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics*, 40, 677-697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E. & Yates, C. J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and distributions*, 17(1), 43-57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>
- Gall, W. K. & Edwards, G. B. (2016) First records for the jumping spiders *Heliophanus kochii* in the Americas and *Myrmarachne formicaria* in New York State (Araneae: Salticidae). *Peckhamia*, 140.1, 1-7.
- Gallien, L., Douzet, R., Pratte, S., Zimmermann, N. E. & Thuiller, W. (2012) Invasive species distribution models—how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography*, 21(11), 1126-1136. <https://doi.org/10.1111/j.1466-8238.2012.00768.x>
- GBIF.org (13 January 2024) GBIF *Myrmarachne formicaria* de Geer Occurrence Download. <https://doi.org/10.15468/dl.79jac6>
- GBIF.org (06 August 2024) GBIF *Formica* L. Occurrence Download. <https://doi.org/10.15468/dl.u4g7kp>
- Giulian, J. Jones, T. C. & Moore, D. (2024) Assessing the potential invasive range of *Trichonephila clavata* using species distribution models. *Journal of Asia-Pacific Biodiversity*, 17(3), 490-496. <https://doi.org/10.1016/j.japb.2024.02.003>
- Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C. & Kueffer, C. (2014) Unifying niche shift studies: insights from biological invasions. *Trends in Ecology & Evolution*, 29(5), 260-269. <https://doi.org/10.1016/j.tree.2014.02.009>
- Guisan A. & Thuiller W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993-1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>
- Hernandez, P. A., Graham, C. H., Master, L. L. & Albert, D. L. (2006) The effect of sample size and species characteristics on performance of different

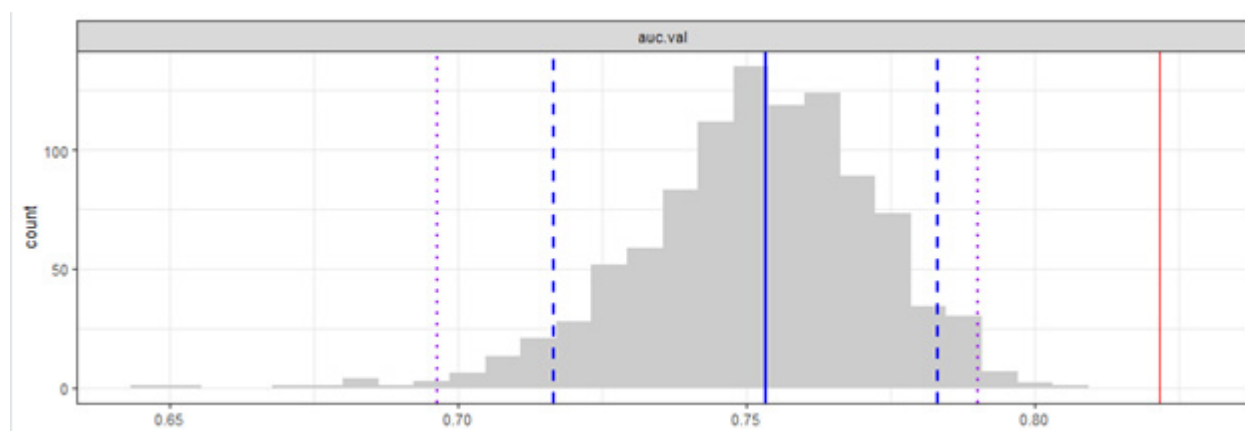
- species distribution modeling methods. *Ecography*, 29(5), 773-785. <https://doi.org/10.1111/j.0906-7590.2006.04700.x>
- Ivanova, N. V. & Shashkov, M. P. (2021) The possibilities of GBIF data use in ecological research. *Russian Journal of Ecology*, 52, 1-8. 10.1134/S1067413621010069
- Kim, S. T. & Lee, S. Y. (2014) Arthropoda: Arachnida: Araneae: Clubionidae, Corinnidae, Salticidae, Segestridae. Spiders. *Invertebrate Fauna of Korea* 21(31), 1-186.
- Lindström, L. Alatalo, R. V. & Mappes, J. (1997) Imperfect Batesian mimicry—the effects of the frequency and the distastefulness of the model. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1379), 149-153.
- Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., et al. (2014) ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in ecology and evolution*, 5(11), 1198-1205. <https://doi.org/10.1111/2041-210X.12261>
- N.A., A., Shaanker, M. U., Bhat H.N., P., Charles, B., Shaanker R., U. & Shah, M. A. (2022) Niche shift in invasive species: is it a case of “home away from home” or finding a “new home”? *Biodiversity and Conservation*, 31(11), 2625-2638. <https://doi.org/10.1007/s10531-022-02447-0>
- Naimi, B., Hamm, N. A., Groen, T. A., Skidmore, A. K. & Toxopeus, A. G. (2014) Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 37(2), 191-203. <https://doi.org/10.1111/j.1600-0587.2013.00205.x>
- Nentwig, W. (2015) Introduction, establishment rate, pathways and impact of spiders alien to Europe. *Biological Invasions*, 17, 2757-2778. 10.1007/s10530-015-0912-5
- O'Donnel, M. S. & Ignizio, D. A. (2012) Bioclimatic predictors for supporting ecological applications in the conterminous United States (No. 691). US Geological Survey.
- Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. & Blair, M. E. (2017) Opening the black box: An open-source release of Maxent. *Ecography*, 40(7), 887-893. <https://doi.org/10.1111/ecog.03049>
- Phillips, S. J., Anderson, R. P. & Schapire, R. E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological modelling*, 190(3-4), 231-259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Pekár, S. & Jiroš, P. (2011) Do ant mimics imitate cuticular hydrocarbons of their models? *Animal Behaviour*, 82(5), 1193-1199. <https://doi.org/10.1016/j.anbehav.2011.09.007>
- Quarrell, S. R., Arabi, J., Suwalski, A., Veuille, M., Wirth, T. & Allen, G. R. (2018) The invasion biology of the invasive earwig, *Forficula auricularia* in Australasian ecosystems. *Biological Invasions*, 20, 1553-1565. <https://doi.org/10.1007/s10530-017-1646-3>
- Shamble P.S., Hoy R.R., Cohen I. & Beatus T. (2017) Walking like an ant: a quantitative and experimental approach to understanding locomotor mimicry in the jumping spider *Myrmarachne formicaria*. *Proceedings of the Royal Society B*, 284: 20170308. <https://doi.org/10.1098/rspb.2017.0308>
- Taucare-Ríos, A., Bizama, G. & Bustamante, R. O. (2016) Using global and regional species distribution models (SDM) to infer the invasive stage of *Latrodectus geometricus* (Araneae: Theridiidae) in the Americas. *Environmental Entomology*, 45(6), 1379-1385. <https://doi.org/10.1093/ee/nvw118>
- Taucare-Ríos, A., Nentwig, W., Bizama, G. & Bustamante, R. O. (2018) Matching global and regional distribution models of the recluse spider *Loxosceles rufescens*: to what extent do these reflect niche conservatism?. *Medical and Veterinary Entomology*, 32(4), 490-496. <https://doi.org/10.1111/mve.12311>
- Warren, D. L. & Seifert, S. N. (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological applications*, 21(2), 335-342. <https://doi.org/10.1890/10-1171.1>
- World Spider Catalog (2023) World Spider Catalog. Version 24.5. Natural History Museum Bern, online at <http://wsc.nmbe.ch>, accessed on: 21 Dec. 2023.
- Wyckhuys, K. A., Kondo, T., Herrera, B. V., Miller, D. R., Naranjo, N. & Hyman, G. (2013) Invasion of exotic arthropods in South America's biodiversity hotspots and agro-production systems. In *Potential invasive pests of agricultural crops* (pp. 373-400). Wallingford UK: CABI.
- Vetaas, O. R. (2002) Realized and potential climate niches: a comparison of four *Rhododendron* tree species. *Journal of Biogeography*, 29(4), 545-554. <https://doi.org/10.1046/j.1365-2699.2002.00694.x>
- Vink, C. J. & Dupérré, N. (2011) *Nesticus eremita* (Araneae: Nesticidae): redescription of a potentially invasive European spider found in New Zealand. *The Journal of Arachnology*, 39(3), 511-514. <https://doi.org/10.1636/A11-57.1>
- Vink, C. J. & Kean, J. M. (2013) PCR gut analysis reveals that *Tenuiphantes tenuis* (Araneae: Linyphiidae) is a potentially significant predator of Argentine stem weevil, *Listronotus bonariensis* (Coleoptera: Curculionidae), in New Zealand pastures. *New Zealand journal of zoology*, 40(4), 304-313. <https://doi.org/10.1080/03014223.2013.794847>



Supplemental Figure 1. Distribution of null AUC values for 1,000 null SDMs built using the global SDM workflow. Dotted purple lines represent the 1st and 99th percentiles; dashed blue lines represent the 5th and 95th percentiles; and the solid blue line represents the 50th percentile. The red line is the AUC value for the global SDM.



Supplemental Figure 2. Distribution of null AUC values for 1,000 null SDMs built using the invasive SDM workflow. Dotted purple lines represent the 1st and 99th percentiles; dashed blue lines represent the 5th and 95th percentiles; and the solid blue line represents the 50th percentile. The red line is the AUC value for the global SDM.



Supplemental Figure 3. Distribution of null AUC values for 1,000 null SDMs built using the native SDM workflow. Dotted purple lines represent the 1st and 99th percentiles; dashed blue lines represent the 5th and 95th percentiles; and the solid blue line represents the 50th percentile. The red line is the AUC value for the global SDM.

Supplemental Table 1. Table of statistics for global, invasive and native SDMs.

| SDM | Features | rm | AICc | AUC (train) | AUC (test) | Null AUC | Null AUC SD | Z-score | P- value |
|------------|-----------------|-----------|-------------|------------------------|-----------------------|-----------------|------------------------|----------------|-----------------|
| Global | LQH | 5 | 9598.37 | 0.8601 | 0.8325 | 0.5868 | 0.03768 | 6.52 | 3.52E-11 |
| Invasive | LQ | 1 | 2878.18 | 0.9464 | 0.9378 | 0.5877 | 0.04987 | 7.02 | 1.11E-12 |
| Native | LQH | 3 | 5804.85 | 0.8575 | 0.8217 | 0.7518 | 0.0153 | 4.57 | 2.47E-06 |