Predicting the occurrence of an endangered salamander in a highly urbanized landscape

Andie L. Siemens[1],*, **Jim P. Bogart[1], Jessica E. Linton[2], D. Ryan Norris[1]**

[1]Department of Integrative Biology, University of Guelph, 50 Stone Rd. E, Guelph, ON N1G 2W1, Canada
[2]Natural Resource Solutions Inc., 415 Phillip St., Waterloo, ON N2L 3X2, Canada

ABSTRACT: Effective protection of threatened species living in highly urbanized landscapes requires detailed information on their population distribution. For species that are difficult to detect, species distribution models (SDMs) can be valuable tools for predicting their occurrence. We created an SDM to predict breeding pond locations of the endangered Jefferson salamander *Ambystoma jeffersonianum* in southern Ontario, the most highly developed and populated region in Canada. Using a maximum entropy modelling algorithm (Maxent), we combined known breeding pond occurrences with climate, land type, soil, and topography data to capture the ecological niche of the Jefferson salamander. Our SDM showed excellent performance (AUC = 0.919), with land type being the most important predictor variable. We produced a continuous habitat suitability map that predicted most hotspots of suitable habitat to occur along the Niagara Escarpment, with small patches in hedge rows and forest fragments in surrounding agricultural areas. Our refined presence–absence map predicted a high suitability area of 305 km^2 with high specificity and moderate overall accuracy. Over half of this area was within the Ontario Greenbelt, demonstrating the importance of protecting this land from future development. Our work demonstrates how SDMs can be used to inform decisions on endangered species and direct conservation efforts towards critical habitats.

KEY WORDS: *Ambystoma jeffersonianum* · Conservation · Ecological niche · Maxent · Ontario Greenbelt · Jefferson salamander · Species distribution modeling

1. INTRODUCTION

Urban and agricultural development poses a major threat to endangered species (Petranka 1998, Porej et al. 2004, McCune 2016, Stevens & Conway 2020). Roads built through habitats can become barriers to dispersal between populations (Compton et al. 2007), isolate populations through habitat fragmentation (Guerry & Hunter 2002, Rothermel & Semlitsch 2006), cause vehicular casualties (Ashley & Robinson 1996), and introduce invasive species that degrade the quality of remaining patches (COSEWIC 2010). Urban and agricultural development may also be sources of pollution (Guerry & Hunter 2002), degrading nearby forest and wetland habitats that are essential to endangered species (Stevens & Conway 2020). To better understand how such threats are affecting endangered species, a critical first step is to estimate population distribution. Knowledge of how species are spatially distributed within their range, and what factors influence their distribution, can inform where to limit development (Porej et al. 2004, van Drunen et al. 2020), prioritize additional surveys and data collection (Rosner-Katz et al. 2020), and direct conservation and management actions (Srivastava et al. 2019, Frans et al. 2022).

Species distribution models (SDMs) are tools that can be used to infer the population distribution of rare species that are difficult to measure directly (Rosner-Katz et al. 2020). SDMs function by correlating the locality data of known occurrences with environmental factors to model the ecological niche of

the species (Lunghi et al. 2018, Mills et al. 2020, Nottingham & Pelletier 2021). Specifically, SDMs estimate the probability of species occurrence at a given location or some scalar of this probability (Rosner-Katz et al. 2020). Projecting the spatial model visualizes the predicted locations of a species and allows for inferences of population distribution (Elith et al. 2020). While SDMs have demonstrated high accuracy in predicting species occurrence, they may be poor in predicting abundance or more advanced ecological parameters, such as population growth rate or diversity (Lee-Yaw et al. 2022). Furthermore, highly suitable habitats predicted by SDMs may not be occupied by a species due to other factors not captured in the SDM creation, usually biotic interactions such as predation and competition (Lee-Yaw et al. 2022). Therefore, population distribution estimated from SDMs should be treated as inferences until validated by ground surveys. Despite these limitations, SDMs remain valuable tools for predicting the population distribution of rare species with limited data from known populations (Srivastava et al. 2019).

The Jefferson salamander *Ambystoma jeffersonianum* is an example of an endangered species with limited information on population distribution (COSEWIC 2010). Ranging across the northeastern USA from the New England region to Indiana, Jefferson salamanders are primarily found in deciduous forest containing fishless ponds suitable for breeding (Petranka 1998). The northern limit of their range extends into Ontario, Canada, where they are found primarily along the Niagara Escarpment in isolated populations, typically with less than 200 adult individuals (COSEWIC 2010). The density of human development in southern Ontario and abundance of roads along the Niagara Escarpment have heavily fragmented forests (Rosner-Katz et al. 2020), which are essential for Jefferson salamanders. As of 2010, of the 87 known historical population locations, only 33 were confirmed to persist (COSEWIC 2010). This apparent decline, in combination with the species' limited dispersal capacity and high fidelity to breeding ponds, has led Jefferson salamanders to be listed as endangered at the provincial (Government of Ontario 2023) and federal (Government of Canada 2023) levels.

In Canada, Jefferson salamanders are included in a complex with blue-spotted salamanders *A. laterale* and unisexual dependents. Unisexuals with 1 *A. laterale* genome and 2 or more Jefferson genomes (e.g. LJJ, LJJJ) require sperm from Jefferson salamanders for recruitment (COSEWIC 2016). These all-female unisexuals have a unique life strategy as obligate sexual parasites; they require the sperm of a male *Ambystoma*, yet their offspring are all-female clones and not hybrids. Like the sexual Jefferson salamander, these unisexual dependents are also classified as endangered in Ontario (Government of Canada 2023) and have a recommended status of endangered in Canada (COSEWIC 2016). Although these unisexuals are dependent on Jefferson salamander populations to reproduce, they have been found to outnumber sexual Jefferson salamanders in shared breeding ponds and are even found in ponds without sexual Jefferson salamanders (Bogart & Klemens 1997, Bogart et al. 2017). Because there is evidence of niche partitioning between sexual Jefferson salamanders and unisexual dependents (van Drunen et al. 2020), this study only models the ecological niche of sexual Jefferson salamanders.

Although the northern extent of the Jefferson salamander's range has been established, there are limited data on the distribution of their populations within the range. As amphibians, Jefferson salamanders spend most of their time hidden under leaf litter to avoid desiccation and escape freezing temperatures by overwintering in underground burrows (Petranka 1998, Rothermel & Semlitsch 2006). This presents a challenge for population surveys, as it is time and resource intensive to observe or capture Jefferson salamanders (Rothermel & Semlitsch 2006, Environment Canada 2016). Surveying ponds used by Jefferson salamanders for breeding is the most feasible way to collect population distribution data. However, data on the true extent of these breeding pond locations are still scarce because the period for surveying is narrow (late March–late April, depending on latitude; Petranka 1998, Rothermel & Semlitsch 2006), and several permits and authorizations are required to conduct surveys. To minimize the risk of desiccation, Jefferson salamanders may only migrate from overwintering sites to breeding ponds on a few rainy nights in early spring and visit the pond for only a few days (COSEWIC 2010). In addition, many of these breeding ponds are seasonal vernal pools that only contain water in the spring, and adults may not breed every year (Petranka 1998, Compton et al. 2007, COSEWIC 2010). Therefore, inferences from breeding pond surveys may underestimate or misrepresent the true population distribution (Fourcade et al. 2014). Fortunately, using SDMs to combine locality data from ground surveys of rare species with data on environmental conditions has been successful in improving population distribution inferences in other species (McCune 2016, Rosner-Katz et al. 2020, Zhang et al. 2020).

To address the knowledge gap regarding the population distribution of Jefferson salamanders, the primary objective of this study was to create an SDM to predict the distribution of Jefferson salamander breeding ponds in southern Ontario. From this SDM, we created 2 maps for 2 distinct purposes: a habitat suitability map for informing conservation planning and a refined presence–absence map for directing survey efforts in identifying new breeding ponds. In doing so, the relative influence of environmental variables in predicting the habitat suitability of these breeding ponds was also assessed. Jefferson salamander breeding pond distribution was modeled by combining known occurrence data of Jefferson salamander breeding ponds from 1979 to 2019 with environmental data hypothesized to be relevant to the ecological niche of Jefferson salamanders (Table 1). A continuous predictive surface of habitat suitability for breeding ponds was created using the maximum entropy (Maxent) modeling algorithm and validated using the area under the receiver-operator curve (AUC score). Binary maps of presence and absence were then created by thresholding the habitat suit-

ability surface, with locations above the threshold classified as present and those below as absent. Sensitivity, specificity, and the true skills statistic (TSS) were used to validate presence–absence maps.

2. METHODS

2.1. Occurrence data

Occurrence data of Jefferson salamander breeding ponds in southern Ontario were obtained from field surveys of Jefferson salamander breeding ponds from 1979 to 2019 compiled by J. P. Bogart. The UTM coordinates of occurrences (n = 173) were plotted in ArcGIS Pro v.2.5 software (ESRI 2020) and cleaned to remove any inaccurate or duplicate points. Any occurrence points that could not be confirmed as a known breeding pond upon re-inspection by J. P. Bogart and J. E. Linton were removed. To avoid duplicates, only the most recent occurrence point was retained for ponds that were surveyed multiple times, leaving a final sample size of 111 known

Table 1. Hypothesized mechanism of environmental variables in predicting distribution of Jefferson salamander breeding ponds

Variable	Hypothesized mechanism	References
Temperature	As ectotherms, Jefferson salamanders are temperature dependent. Jefferson salamanders are more active at the surface at higher temperatures, spending less time sheltering and more time foraging for prey. Alternatively, lower temperatures may increase surface moisture by reducing evaporation, allowing for increased surface activity without risk of desiccation.	COSEWIC (2010), Peterman & Semlitsch (2013), Lunghi et al. (2018), Mills et al. (2020)
	Additionally, temperature may affect the survival of Jefferson salamanders in early aquatic life stages. Egg and larval development may be affected by water temperature, specifically the length of the embryonic and larval periods. Large temperature fluctuations in spring may also cause breeding ponds to freeze or water levels to fall via evaporation, leading to freezing or desiccation of eggs.	
Precipitation	As amphibians, Jefferson salamanders are at risk of desiccation and anoxia. in dry environments. Therefore, moist habitats with high precipitation levels may be favoured. Additionally, precipitation may be a source of inundation for vernal ponds used for breeding.	Keeley & Zedler (1998), COSEWIC (2010), Nottingham & Pelletier (2021)
Topography	Topography, specifically the depth of depressions and the slope of the surrounding ground, may affect filling of vernal pools in spring and maintaining water levels into summer, thus affecting the ability of Jefferson salamanders to complete metamorphosis.	McCune (2016), Mills et al. (2020)
Soil type	Soil type may affect water drainage and infiltration in vernal ponds, thereby affecting their ability to maintain water levels long enough to complete metamorphosis.	Keeley & Zedler (1998), Rains et al. (2008)
Land type	Forest canopy cover may affect surface moisture by blocking solar radiation and limiting evaporation, thereby influencing surface activity and foraging behaviour. Intact forested area may affect habitat suitability of breeding ponds by minimizing movement barriers between ponds and over-wintering sites.	Guerry & Hunter (2002), Peterman & Semlitsch (2013)

breeding pond occurrences (Fig. 1). Using historical occurrence records introduces a concern that some sites may no longer be actively used for breeding or that some ponds no longer exist. Many historical sites have been re-visited by Jefferson salamander recovery team members, and their findings have been included in the most recent COSEWIC update. We were careful to use the most recently updated occurrence records in constructing the SDM.

We are confident that there was little to no spatial bias in the occurrence records we used to construct the SDM. Because the Jefferson salamander is part of a complex of sexual and unisexual salamanders, genetic testing is required to identify members of the complex found during surveys. Thus, genetic testing has been conducted on many salamander surveys across their Ontario range and would have documented Jefferson salamanders if they were present. Additionally, blue-spotted salamanders are part of this complex and have a much wider distribution in Canada than the Jefferson salamander. Therefore, surveys for the blue-spotted salamander would also identify Jefferson salamanders if they were present.

2.2. Environmental data

To construct the SDM, a set of environmental variables hypothesized to be relevant to the ecological niche of Jefferson salamanders were selected as predictor variables (Table S1 in the Supplement at www.int-res.com/articles/suppl/n052p081_supp.pdf). To capture the fundamental niche of Jefferson salamanders, climatic data sourced from WorldClim at 1 km spatial resolution were selected (Trumbo et al. 2012, Pelletier et al. 2015, Nottingham & Pelletier 2021). These data consisted of 19 variables representing annual trends of precipitation and temperature (BIO 1–19) derived from monthly temperature and rainfall records from 1970 to 2000. However, climatic variables may show little variation at the scale of the Jefferson salamander range in southern Ontario, specifically the Niagara Escarpment. Therefore, environmental variables capable of capturing habitat variation were also included to improve the ability of the SDM to predict breeding pond distribution at a local scale (Trumbo et al. 2012, Koma et al. 2022). To model how environmental variables influence distribution at a local scale, topography, soil type, and land type were also included as predictor variables. To capture the effect of topography on filling and maintaining potential breeding ponds, topographic data at 30 m spatial resolution were obtained

from the imagery-derived Ontario Provincial Digital Elevation Model (PDEM; geohub.lio.gov.on.ca). In addition to raw elevation, a second topographic variable termed 'probability of depression' was calculated using WhiteBoxTools (Lindsay 2016), software that uses elevation, slope, and aspect from an input digital elevation model to calculate the probability of a location being in a terrain depression. To capture the influence of soil type on water drainage in potential breeding ponds, data on soil type were obtained from the Soil Survey Complex (v.4) dataset from Land Information Ontario (geohub.lio.gov.on.ca). From this dataset, categorical variables of soil texture class and soil drainage class were extracted at 30 m spatial resolution. To capture the effect of land type on the habitat suitability of breeding ponds, data on land type from the Southern Ontario Land Resource Information System (SOLRIS v.3) were obtained at 15 m spatial resolution (geohub.lio.gov.on.ca).

2.3. Model building

The Maxent algorithm was chosen to fit the SDM due to the algorithm's popularity in ecological studies using presence-only data and its high performance even with small sample sizes (Fourcade et al. 2014, McCune 2016, Elith et al. 2020, Mills et al. 2020, Nottingham & Pelletier 2021). In a general sense, the Maxent algorithm predicts a geographic distribution at maximum entropy (closest to uniform) while being constrained by environmental conditions learned at the occurrence points (Phillips et al. 2017). To address the lack of true absence data, Maxent generates an array of randomly distributed background points (n = 10 000) and compares the environmental conditions at these points to those at occurrence points (Merow et al. 2013). The resultant model predicts habitat suitability on a scale of 0 to 1, which can be projected onto a map as a continuous predictive surface. A binary map of predicted presence and absence can then be produced by thresholding habitat suitability. Values above the threshold are considered to indicate presence, while those below indicate absence.

To build an SDM using Maxent, all input environmental data are required to be in raster grids of the same spatial resolution, projection, and extent. As such, all environmental data were aligned in ArcGIS Pro prior to model training. It is recommended to train the Maxent model using background points from the same geographic distribution as the occurrence points to avoid including novel environmental
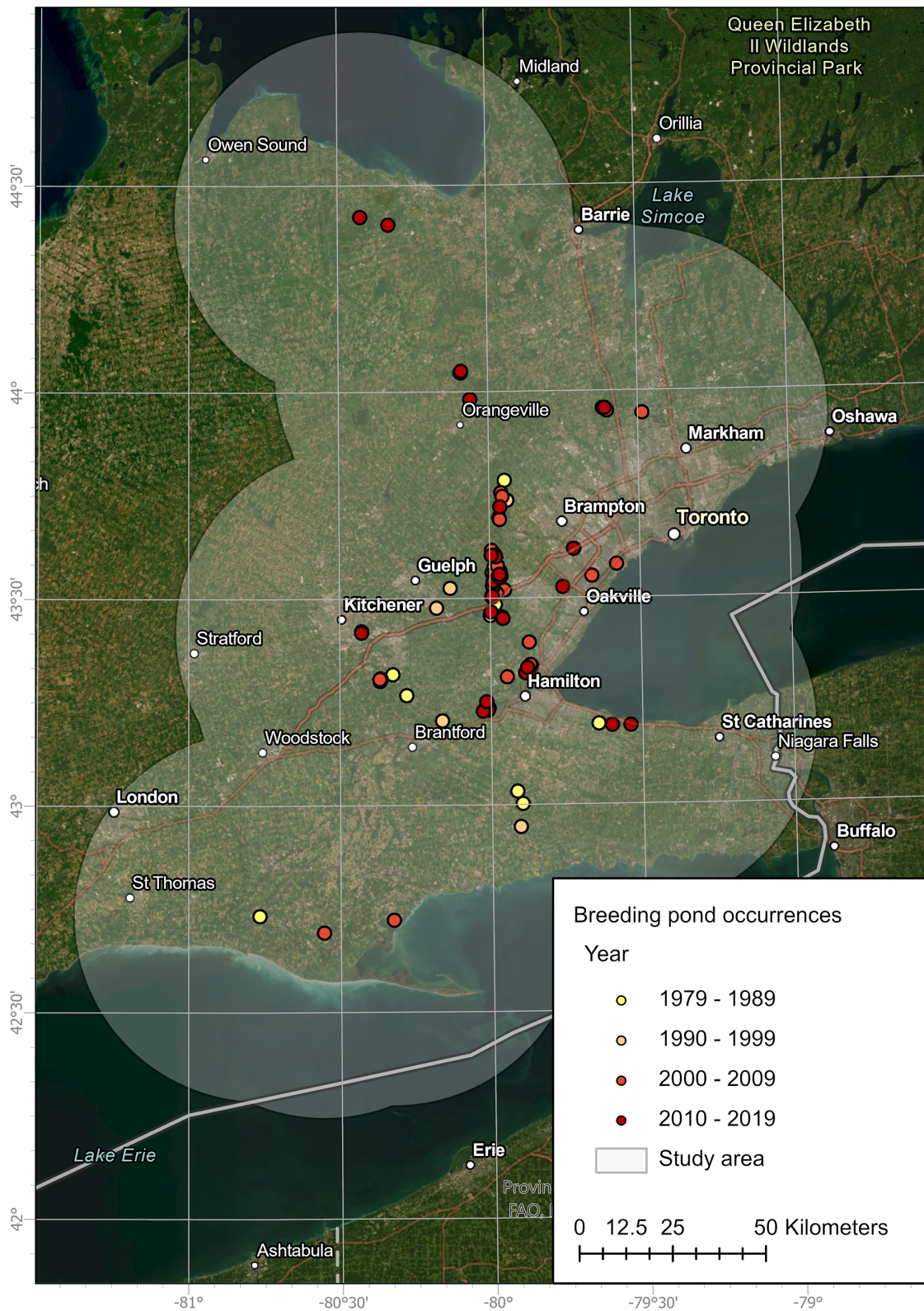
Fig. 1. Occurrences of Jefferson salamander breeding ponds (n = 111) in southern Ontario, categorized by most recent year recorded. Most recent points are along the Niagara Escarpment between Guelph and the Greater Toronto area. The study area was defined by a 50 km buffer around all occurrence points. Mapped using NAD83 UTM Zone 17N projection in ArcGIS Pro

conditions never encountered by the species (Phillips 2008, Jarnevich & Young 2015). Therefore, a study area was defined by buffering the occurrence points by 50 km and was used to restrict the geographic extent of the model (Fig. 1). All environmental raster grids were clipped to the study area, forcing Maxent to generate random background points from within the study area boundary.

Maxent employs machine learning techniques to learn the environmental constraints that influence habitat suitability. Therefore, there can be variation between model runs even when using identical input data and parameter settings caused by the inherent stochasticity of machine learning (Merow et al. 2013, Jarnevich & Young 2015). To account for this random variation, the model was run with 10 replicates, and the results were averaged to create the final SDM. The same occurrence points, background points, and environmental variables were used for each replicate. However, a random seed was used to ensure a different 80% of the occurrence points were used to train the model, with the remaining 20% used to test the model's fit. The complementary log-log transformation was used as the model output instead of the more common logistic transformation, as recommended by the developers of Maxent's original (Phillips et al. 2006) and open-source (Phillips et al. 2017) algorithms.

2.4. Model calibration

Statistical analyses were performed in ENMTools (Warren et al. 2010) and R software v.4.1.1 (R Core Team 2021). Although Maxent is generally robust to the collinearity of environmental predictor variables (Elith et al. 2011, Feng et al. 2019), reducing correlation between variables creates a more parsimonious model with interpretable variable response curves (Merow et al. 2013). Given that one of the study objectives was to interpret the relative influence of environmental predictor variables, Pearson's correlation coefficient was calculated for all continuous variables in ENMTools (Warren et al. 2010; Fig. S1). Variable pairs with a correlation coefficient greater than |0.7| (Feng et al. 2019) were identified, and the variable with the lower percent contribution to the model was removed.

To control for overfitting, Maxent's beta regularization multiplier can be calibrated to balance model complexity with model fit and prevent the model from describing random noise in the training data (Warren & Seifert 2011, Morales et al. 2017). The beta regularization multiplier penalizes model complexity by only retaining coefficients that improve model fit above the beta value and removing all others. Following a method adapted from Jarnevich & Young (2015), 10 models were run, each with 10 replicates, while varying the beta value from 1 to 10. To allow for comparisons between the models, the same background points were used for all model runs. For the first model run, the coordinates of the automatically generated background points were recorded to create species with data (SWD) files consisting of environmental data for each variable at each coordinate point. These SWD files provide another way to supply Maxent with input occurrence and environmental data and allow the user to specify the background points to be used in model training. After running all 10 models with varying beta values, the corrected Akaike's information criterion (AICc) scores were calculated using ENMTools (Warren et al. 2010). The beta value corresponding to the smallest AICc score for each replicate was then selected and averaged for an optimized beta value of 4.2. This value was used to run the final calibrated model.

2.5. Model validation

The final calibrated model was validated using the mean AUC score on test data (Trumbo et al. 2012, Rosner-Katz et al. 2020). This curve plots model sensitivity against 1 − specificity, which is calculated as the fractional predicted area in the case of a presence-only model such as this one (Phillips et al. 2006). The mean curve for all 10 replicates was calculated, along with the SD. A mean AUC score below 0.5 indicates a model with a performance worse than random, while a score of 1 indicates perfect model prediction. SD reflects random stochasticity in machine learning because all replicates were run using identical input data and background points.

A jackknife analysis was performed to measure environmental variable importance within the top model. The model was re-run using each variable in isolation while measuring the model gain compared to the original model gain with all variables. To measure the decrease in gain from omitting each variable, this was repeated using all variables but one. Comparing the effect on model gain gives insight on the amount of useful and unique information within each variable for the model (Phillips et al. 2006, 2017).

Sensitivity, specificity, and TSS were calculated for both presence−absence maps to validate their accu-

racy and support their application to surveying for Jefferson salamander breeding ponds. TSS was chosen because it mathematically outperforms kappa, another commonly used statistical measure of model performance, and is independent of species prevalence (Allouche et al. 2006, Liu et al. 2013). All 3 measures were calculated for each replicate and averaged for final values.

2.6. Presence–absence thresholding

To create a binary presence–absence map for directing survey efforts, a threshold value was used to classify the habitat suitability predicted by the SDM as presence or absence. Maxent automatically calculates multiple statistical thresholds that are commonly used to create binary maps from continuous predictive surfaces (Merow et al. 2013). The maximized test sensitivity plus specificity (MaxSSS) threshold has been found to perform equally well using presence-only data or presence–absence data (Liu et al. 2013). This threshold is defined as the value that produces the largest summed value of sensitivity and specificity on the model's test data and therefore produces a higher TSS (sensitivity + specificity − 1) than other thresholding methods (Liu et al. 2013). In this case, sensitivity describes the model's ability to correctly predict true presences and is calculated as 1 − omission error rate using the breeding pond occurrence points. Specificity describes the model's ability to correctly predict true absences and is usually calculated as 1 − commission error rate. However, presence-only models like Maxent lack absence data and therefore cannot calculate the commission error rate (Liu et al. 2013). In such cases, the fraction of the total study area predicted as present can be substituted for the commission error, and this is what Maxent uses to calculate specificity (Phillips et al. 2006).

The MaxSSS threshold was chosen to create the first presence–absence map because it produces a binary map with high overall accuracy by maximizing the TSS score (Allouche et al. 2006). However, this threshold weighs sensitivity and specificity the same in terms of importance. Because one of the research objectives of this study was to direct survey efforts in identifying new breeding ponds, the risk of spending resources to survey ponds incorrectly predicted as present may be more detrimental than the risk of missing ponds incorrectly predicted as absent. Therefore, a second user-defined threshold that prioritizes specificity over sensitivity was selected to create a second, highly specific presence–absence map to be used to direct survey efforts. To select this threshold, the omission error rate and fractional predicted area were plotted against the threshold value to select a value that predicts a small area as present to direct survey efforts while maintaining an acceptable omission rate.

Although a single pixel (30 × 30 m) may be predicted by the model as highly suitable habitat for a breeding pond, Jefferson salamanders also require intact surrounding forest habitat to successfully migrate to and from breeding ponds (Porej et al. 2004, Rothermel & Semlitsch 2006, van Drunen et al. 2020). Therefore, the presence–absence map was further cleaned by removing areas predicted as present that were not surrounded by appropriate habitat. Appropriate habitat was defined as any one of the top suitable land types defined by the model. Using the Focal Statistics tool in ArcGIS Pro, a moving window of 3 × 3 pixels (90 × 90 m) was used to calculate the number of pixels in each neighbourhood that were classified as one of these land types. Only pixels previously predicted as present by the model that also were located on, and surrounded by, appropriate habitat remained classified as present. All other pixels were reclassified as absent.

3. RESULTS

3.1. Habitat suitability

The SDM predicted a continuous surface of habitat suitability covering an extent of 29 206 km², reaching from the southern tip of Ontario to Georgian Bay (Fig. 2a). Validated using the withheld 20 % of occurrence points and averaged across all 10 model replicates, the final calibrated SDM showed excellent performance, with a mean AUC score of 0.919 (Fig. S2). The SDM predicted highly suitable habitat (>0.9) along the Niagara Escarpment, with hotspots in the Halton Falls Conservation Area, Crawford Lake Conservation Area, and Dundas Valley Conservation Area and along a narrow stretch of forest east of Erin. Land surrounding the escarpment displayed a striped pattern, with suitable (>0.6) forest patches and hedge rows separated by unsuitable (<0.3) agriculture land. Most cities and urbanized areas were predicted as highly unsuitable (<0.1). However, urban areas such as Toronto, Mississauga, Oakville, Burlington, and Hamilton along the north shore of Lake Ontario showed moderate habitat suitability (0.2–0.7).
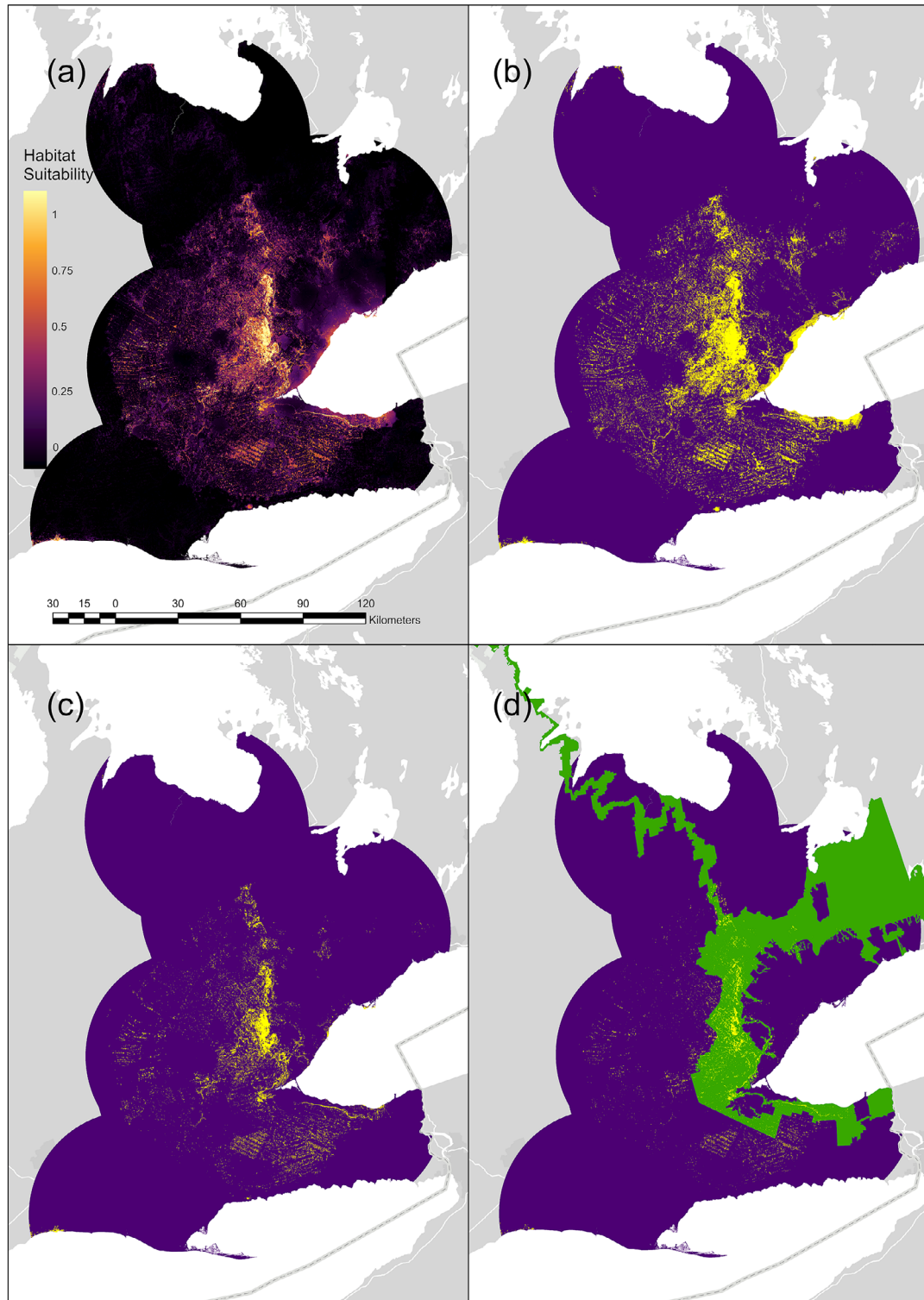
Fig. 2. (a) Species distribution model predicting the habitat suitability of breeding ponds created using the maximum entropy algorithm along occurrence data and environmental variables (climate, land type, soil type, and topography). Habitat suitability increases with colour warmth (dark purple to bright yellow). Binary presence–absence maps were created by thresholding habitat suitability at (b) maximum sensitivity plus specificity and (c) a user-defined threshold of 0.626 for high specificity. Yellow areas indicate presence; purple areas indicate absence. In (d), the high-specificity map was refined by removing predicted presences not surrounded by appropriate forest habitat to create a map for directing survey efforts in identifying new breeding ponds. The Ontario Greenbelt designation (Scholars GeoPortal) is shown in green

3.2. Relative influence of environmental predictor variables

The top 4 contributing variables to the model were land type (42%), mean temperature of the wettest quarter (23%), soil texture (16%), and temperature seasonality (11%) (Table 2). To evaluate the relationships between the environmental variables and habitat suitability, the response curves for each variable were plotted with all other variables held constant at their respective means. The land types associated with the highest habitat suitability (>0.9) were thicket swamp, deciduous forest, mixed forest, and treed swamp (Fig. 3a). The least suitable land types were roads (0.68), undifferentiated land (0.59), and agriculture fields (0.29). For soil textures, loam and sandy clay loam were associated with higher habitat suitability (>0.9) than other soil textures (Fig. 3b). All soil textures were associated with habitat suitability

Table 2. Relative contribution of environmental predictor variables to species distribution model (SDM) of Jefferson salamander breeding ponds in southern Ontario. Relative contribution was calculated as the increase or decrease in regularized gain per iteration of the Maxent training algorithm, averaged across 10 model replicates

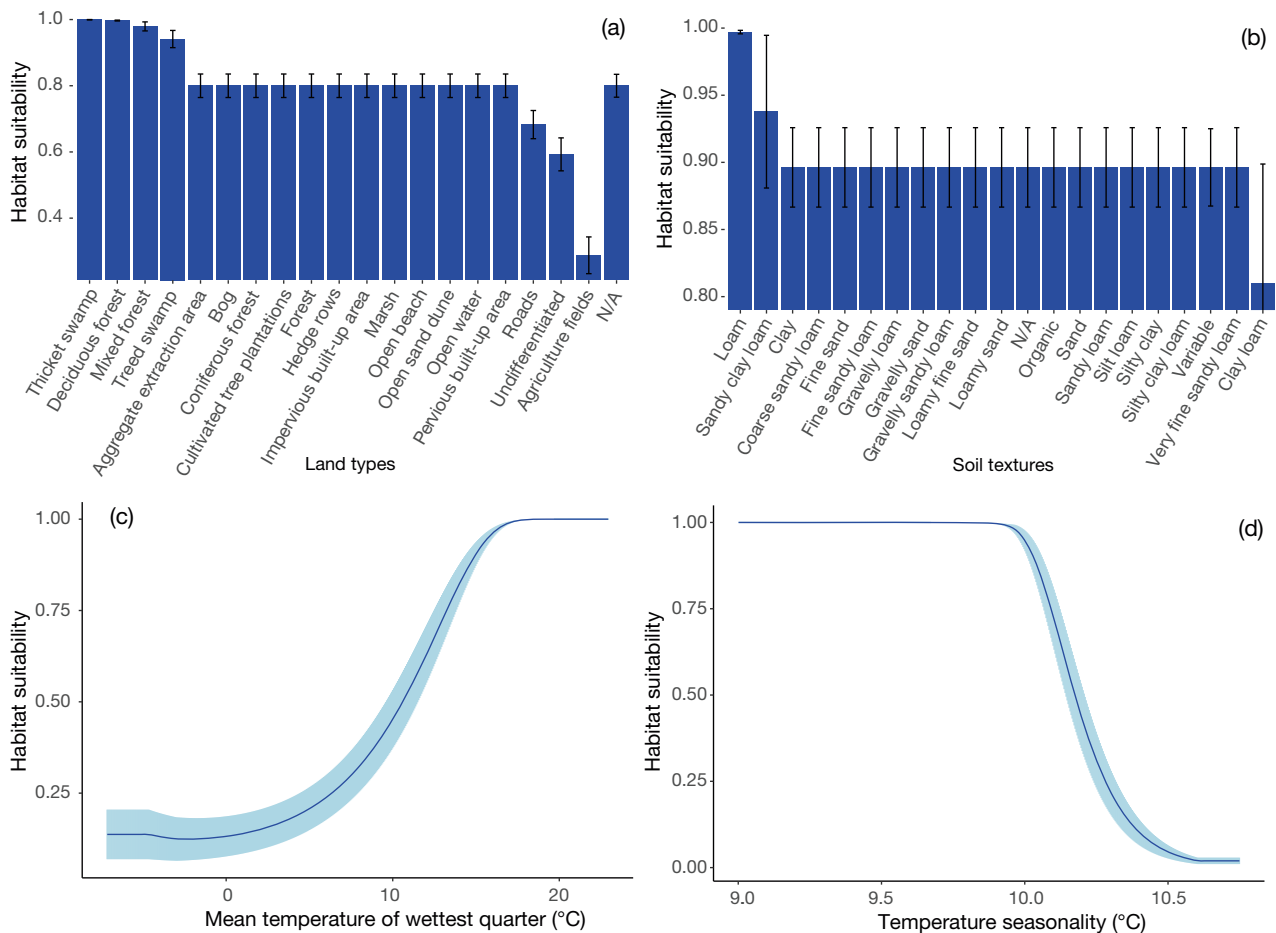| Environmental variable	Contribution to SDM (%)
Land type	42.4
Mean temperature of wettest quarter	22.8
Soil texture	15.6
Temperature seasonality	10.5
Isothermality	4.0
Soil drainage class	2.6
Mean temperature of warmest quarter	1.0
Precipitation of warmest quarter	0.4
Annual precipitation	0.4
Mean temperature of driest quarter	0.2
Precipitation seasonality	0.1
Probability of terrain depression	0

Fig. 3. Response curves of top contributing environmental variables to the species distribution model of Jefferson salamander breeding ponds in southern Ontario. Responses averaged across 10 model replicates (dark blue lines and bars), with all other variables held constant at their respective means. (a,b) Black error bars and (c,d) light blue area represent ±1 SD between model replications. Note that the scale of the habitat suitability axis may differ between figures

Fig. 4. Jackknife analysis of environmental variable importance for the species distribution model (SDM) of Jefferson salamander breeding ponds in southern Ontario. Green bars show model gain on test data while omitting each variable. Blue bars show model gain while using each variable in isolation. Red bars show the total gain of the SDM using all variables. Land type had the highest gain when used in isolation and the largest decrease in gain when omitted from the model, suggesting it as the most important variable to the SDM. Probability of terrain depression had negative gain when used in isolation, showing that the model performed worse than random with only this variable
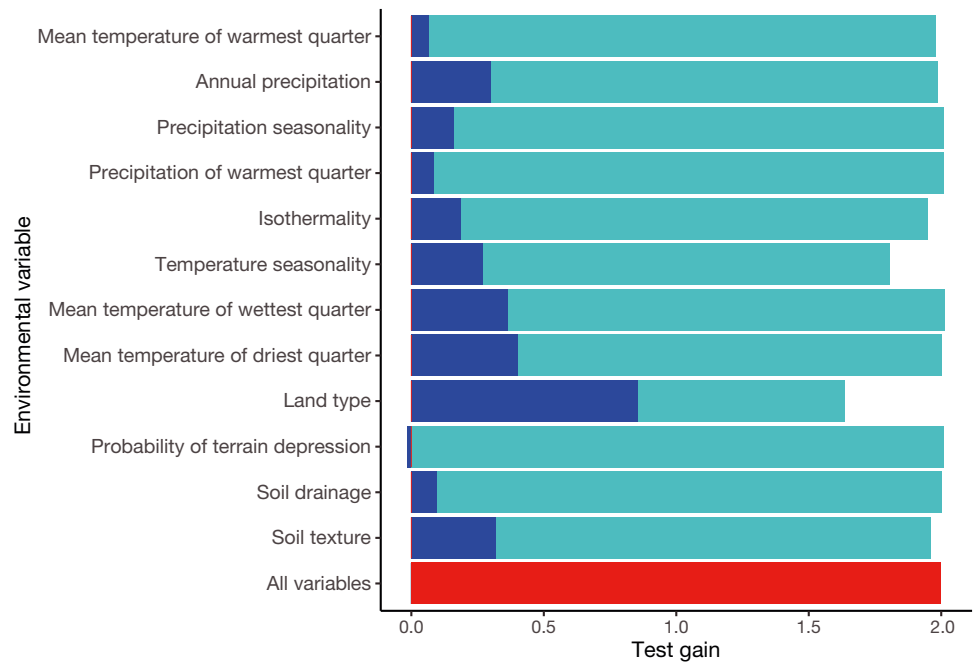
greater than 0.80, and all, except for loam, varied greatly between model replicates. Habitat suitability increased exponentially with mean temperature of the wettest quarter and plateaued above 15°C (Fig. 3c). For temperature seasonality, habitat suitability decreased rapidly from highly suitable (>0.95) to highly unsuitable (<0.05) between 10 and 10.5°C (Fig. 3d). Isothermality and soil drainage class contributed minimally to the model (4 and 2.6%, respectively). Mean temperature of the warmest quarter, precipitation of the warmest quarter, annual precipitation, mean temperature of the driest quarter, precipitation seasonality, and probability of terrain depression each contributed less than 1% to the model.

A jackknife analysis was performed to measure the importance of each environmental variable to the SDM's performance, averaged across model replicates. Land type had the highest gain when used in isolation, suggesting that it had the most useful information for the model on its own (Fig. 4). Land type also had the largest decrease in gain when omitted from the model, suggesting it also had unique information that was not present in the other variables.

3.3. Presence–absence maps

Binary maps of the predicted presence and absence of breeding ponds were produced by thresholding the habitat suitability values from the SDM. Two different thresholds were used to produce presence–absence maps: the MaxSSS threshold and our user-defined threshold. Averaged across all 10 model replicates, the mean MaxSSS threshold for the SDM was 0.316. However, this value was highly variable between model replicates, with an SD of 0.137 (Table 3). Thresholding habitat suitability at 0.316 produced a reasonably accurate presence–absence map (mean TSS = 0.773, mean specificity = 0.896, mean sensitivity = 0.877) that predicted 2331 km^2 of possible breeding pond habitat (Fig. 2b).

The second presence–absence map created with our user-defined threshold produced a highly specific map for directing survey efforts of breeding ponds (Fig. 2c). A threshold of 0.626 was selected by plotting the mean model commission and omission error rate against threshold values and selecting a threshold value with high specificity and acceptable sensitivity (Fig. S3). The resultant presence–absence map was highly specific, at the expense of sensitivity, but still maintained acceptable accuracy (mean specificity = 0.968, mean sensitivity = 0.677, TSS = 0.644) and had smaller SDs between model replicates than with the MaxSSS threshold (Table 3). This map originally predicted 735 km^2 of possible breeding pond habitat, less than a third of the area predicted using the MaxSSS threshold.

Breeding ponds likely require appropriate surrounding habitat to allow for Jefferson salamanders to migrate between ponds and overwintering sites in the forest (Rothermel & Semlitsch 2006). Thus, we

Table 3. Performance of binary presence–absence maps derived from habitat suitability as predicted by the species distribution model using the maximum sensitivity plus specificity (MaxSSS) threshold and our user-defined threshold. Values are shown for all 10 model replicates, along with the means and SDs used to create the maps. True skill statistic (TSS) calculated as sensitivity + specificity − 1

Model replicate	MaxSSS threshold				User-defined threshold for high specificity			
	Threshold	Specificity	Sensitivity	TSS	Threshold	Specificity	Sensitivity	TSS
0	0.246	0.884	0.864	0.748	0.626	0.970	0.682	0.652
1	0.419	0.929	0.909	0.838	0.627	0.964	0.818	0.782
2	0.263	0.892	0.864	0.756	0.625	0.969	0.727	0.696
3	0.292	0.898	0.909	0.807	0.624	0.965	0.682	0.647
4	0.257	0.885	0.909	0.794	0.626	0.967	0.682	0.648
5	0.208	0.871	0.864	0.735	0.624	0.970	0.500	0.470
6	0.207	0.868	0.864	0.732	0.628	0.971	0.545	0.517
7	0.159	0.814	1.000	0.814	0.627	0.966	0.727	0.693
8	0.582	0.962	0.773	0.735	0.629	0.967	0.682	0.649
9	0.530	0.958	0.818	0.776	0.629	0.967	0.727	0.695
Mean	0.316	0.896	0.877	0.773	0.626	0.968	0.677	0.645
SD	0.137	0.042	0.058	0.036	0.002	0.002	0.087	0.085

refined the presence–absence map by removing suitable locations that were not also surrounded by deciduous forest, mixed forest, thicket swamp, and/or treed swamp. This decreased the total area predicted as present from 735 to 305 km^2 (Fig. 2d).

4. DISCUSSION

We created two informative SDMs of Jefferson salamander breeding ponds: a continuous map of habitat suitability for informing conservation planning (Fig. 2a), and a refined presence–absence map for directing survey efforts in identifying new breeding ponds (Fig. 2d). The presence–absence map was designed to be highly specific by thresholding the SDM and was further refined by removing locations not surrounded by appropriate forest habitat to account for migration between ponds and overwintering sites. This map showed patchy distribution of breeding ponds throughout southern Ontario, with high density along the Niagara Escarpment, which agrees with field reports (COSEWIC 2010). We emphasize that this refined presence–absence map was created for the primary purpose of directing survey efforts to identify new breeding ponds, meaning that specificity was prioritized over sensitivity as to minimize resources related to incorrectly predicting a location of a breeding pond in the field. However, as a result, some breeding ponds may have been omitted in this map due to the low sensitivity. Therefore, this presence–absence map is a conservative estimate of breeding pond distribution and may not be appropriate for decision making related to a con-

servation boundary or zoning without further validation. Instead, the continuous habitat suitability map (Fig. 2a) would be better suited for such applications.

Using variable response curves and jackknife analysis, our results suggest that land type was the strongest determinant of breeding pond suitability. As hypothesized, deciduous forest and forested swamp land types that provided appropriate overwintering habitat were associated with high habitat suitability. Due to the limited dispersal range of Jefferson salamanders (van Drunen et al. 2020) and the risk in crossing roads to migrate between breeding ponds to overwintering sites (Guerry & Hunter 2002, Compton et al. 2007, COSEWIC 2010, Environment Canada 2016), it is crucial to identify and protect breeding ponds within forests appropriate for overwintering. Soil texture was another contributing variable to habitat suitability, with loamy soils associated with higher habitat suitability than other soil textures. Although soil texture was the third most important variable contributing to the SDM, all soil types were associated with high habitat suitability (>0.8). This suggests that soil type may be more important in describing habitat quality rather than differentiating between suitable and unsuitable habitat. There is limited research on the effects of soil on Jefferson salamanders, with some studies focusing on soil pH (Horne & Dunson 1994a,b). Studies investigating the effect of other soil characteristics have focused on the red-backed salamander *Plethodon cinereus*, a species from a different family than the Jefferson salamander (Taub 1961, Jaeger 1980, Frisbie & Wyman 1992, Sugalski & Claussen 1997). Therefore, our findings highlight the need for further

research regarding the importance of soil texture in determining Jefferson salamander distribution.

Climate was also important in determining habitat suitability, with temperature variables contributing more to the SDM than precipitation. Habitat suitability increased with mean temperature of the wettest quarter and decreased with seasonality, demonstrating a preference for warm springs and mild seasons. These results agreed with the hypothesis that large temperature fluctuations in spring would decrease habitat suitability by causing breeding ponds to freeze or evaporate, leading to freezing or desiccation of Jefferson salamander adults, egg masses, or larvae. Our results also suggested that habitat suitability increased with spring temperature, which was contrary to previous studies suggesting salamanders are more active at the surface in lower temperatures with less solar exposure (Peterman & Semlitsch 2013, Lunghi et al. 2018). However, as this study modeled breeding ponds and not Jefferson salamanders directly, these results may indicate that warmer spring temperatures thaw breeding ponds faster and allow Jefferson salamanders to migrate earlier in the breeding season.

Contrary to our predictions, the probability of terrain depression did not contribute to the SDM. The probability of terrain depression would be expected to be high at occurrence points because they are known breeding ponds that must be in a terrain depression. Upon further inspection, values for this variable at occurrence points were skewed left, explaining its lack of contribution to the SDM. This may be due to the spatial resolution of the PDEM source data not being fine enough to detect the small shallow depressions for vernal pools. Although elevation data were available at finer resolution, up to 1 m resolution for the LiDAR-derived High Resolution Digital Elevation Model (geogratis.gc.ca), the processing power required to handle such large data sizes was not feasible for this study. Further investigation is needed into the effects of terrain depression on breeding pond distribution by studies with access to greater processing power or by creating higher-resolution SDMs at smaller geographical extents.

We acknowledge that our results are somewhat limited by the spatial resolution of available environmental data. Maxent requires all environmental data to be at the same spatial resolution, which may have introduced errors when the data were resampled to the same resolution. Land type, soil texture, and soil drainage data were resampled from finer resolutions to match elevation data at 30 m resolution. In doing so, some information that may have helped describe

fine-scale patterns of breeding pond distribution was lost (Zhou & Zhang 2014). The finest-resolution climate data available were 1 km, which would produce too coarse an SDM for the purpose of identifying new breeding ponds. Interpolating the climate data to a finer resolution using a bilinear technique undoubtedly introduced errors (Olivero et al. 2016); however, these were thought to be outweighed by the benefit of producing a detailed SDM with the other finer-resolution variables.

A further limitation is that possible biotic interactions were not accounted for in the SDM. Carnivorous fish prey on Jefferson salamanders at all life stages, which may influence the suitability of ponds for breeding (Petranka 1998, Porej et al. 2004, Environment Canada 2016). Similarly, competition with other salamander species and unisexual dependents may affect breeding success for Jefferson salamanders (Bogart et al. 2017, Mills et al. 2020). As data on the distribution of fish and salamander species were not included in building this SDM, the predicted habitat suitability of breeding ponds does not reflect the presence of predatory fish or other salamander species. Further studies that incorporate the biotic interactions between Jefferson salamanders and other species are needed to produce more accurate SDMs of Jefferson salamanders.

Finally, the nature of our data being presence only may underestimate species occurrences and overestimate species absences in our SDM. Occupancy modelling can improve the predictive capabilities of SDMs (Jha et al. 2022). However, this would require a large data size and repeated visits to locations, which was not feasible for this study. Since the protection of the Jefferson salamander in southern Ontario will require ongoing monitoring, future studies should utilize the growing database to incorporate occupancy modelling into SDMs.

Despite limitations involving spatial resolution and the lack of biotic interactions captured by this SDM, our work highlights the importance of protecting Jefferson salamanders in southern Ontario. On the continuous map of habitat suitability, hotspots of highly suitable habitat for breeding ponds were shown in conservation areas surrounded by unsuitable urban and agricultural land (Fig. 2a). The refined presence–absence map predicted roughly 60% of the area to be suitable for breeding ponds (183 of 305 km^2; Fig. 2d) as a conservative estimate. Additionally, this map showed that 78% of known breeding ponds (87 of 111) were within Ontario's Greenbelt, mostly in the Niagara Escarpment Plan designation (100 km^2), followed by the Protected Countryside designation

(81 km^2; Fig. 2d). This highlights that the Greenbelt is of high conservation value for this species. However, this conservation value is threatened by recent policies by the provincial government that will allow further development in the Greenbelt (Jones 2022), potentially destroying breeding ponds or, at the very least, isolating them from other ponds and overwintering sites (Porej et al. 2004, Rothermel & Semlitsch 2006, Environment Canada 2016, Stevens & Conway 2020, van Drunen et al. 2020). Both our habitat suitability map and refined presence–absence map provide important tools for predicting the suitable habitat of an endangered species and directing further survey efforts in the face of future development in the region. Although statistical metrics suggested our SDM performed well, future studies should endeavour to test its accuracy via ground-truthing surveys at the locations predicted as present for Jefferson salamander breeding ponds.

SDMs are valuable tools that support decision making in the conservation and management of endangered species, such as the Jefferson salamander. Particularly for cryptic species that are not easily detected by surveying, SDMs can provide estimates of the spatial distribution of rare species (McCune 2016, Rosner-Katz et al. 2020). By predicting areas of highly suitable habitat, limited resources can be targeted towards protecting critical habitat for these species. Specifically, SDMs can help determine where to limit urban and agricultural development and where to increase habitat protection, improving conservation efficiency (Guisan et al. 2013). This study created a high-performing SDM of Jefferson salamander breeding ponds, from which two maps were produced to inform conservation planning and direct survey efforts in identifying new ponds. In addition to identifying critical habitat in the present, SDMs can also inform how environmental variables may have influenced species distribution in the past and in the future. By projecting SDMs onto historical or future climate models, SDMs provide an opportunity to examine a species' evolutionary history or phylogeny (Pelletier et al. 2015) and predict their potential range shift in response to future environmental changes (Nottingham & Pelletier 2021). Overall, the ability of SDMs to predict a species' spatial distribution and ecological niche in the past, present, and future makes these models invaluable tools for informing conservation strategies for endangered species.

LITERATURE CITED

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J Appl Ecol 43: 1223–1232

Ashley EP, Robinson JT (1996) Road mortality of amphibians, reptiles and other wildlife on the Long Point Causeway, Lake Erie, Ontario. Can Field Nat 100:403–412

Bogart JP, Klemens MW (1997) Hybrids and genetic interactions of mole salamanders (*Ambystoma jeffersonianum* and *A. laterale*) (Amphibia, Caudata) in New York and New England. Am Mus Novit 3218:1–78

Bogart JP, Linton JE, Sandilands A (2017) A population in limbo: unisexual salamanders (genus *Ambystoma*) decline without sperm-donating species. Herpetol Conserv Biol 12:31–55

Compton BW, McGarigal K, Cushman SA, Gamble LR (2007) A resistant-kernel model of connectivity for amphibians that breed in vernal pools. Conserv Biol 21: 788–799

COSEWIC (Committee on the Status of Endangered Wildlife in Canada) (2010) COSEWIC assessment and status report on the Jefferson salamander *Ambystoma jeffersonianum* in Canada. Committee on the Status of Endangered Wildlife in Canada, Ottawa

COSEWIC (2016) COSEWIC assessment and status report on the unisexual *Ambystoma, Ambystoma laterale*, small-mouthed salamander-dependent population, Jefferson salamander-dependent population and the blue-spotted salamander-dependent population, in Canada. Committee on the Status of Endangered Wildlife in Canada, Ottawa

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. Divers Distrib 17:43–57

Elith J, Graham C, Valavi R, Abegg M and others (2020) Presence-only and presence-absence data for comparing species distribution modeling methods. Biodivers Inform 15:69–80

Environment Canada (2016) Recovery strategy for the Jefferson salamander (*Ambystoma jeffersonianum*) in Canada. Species at Risk Act (SARA) Recovery Strategy Series. Environment Canada, Ottawa

Esri (2022) ArcGIS Pro (Version 3.0). Esri Inc. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview

Feng X, Park DS, Liang Y, Pandey R, Papeş M (2019) Collinearity in ecological niche modeling: confusions and challenges. Ecol Evol 9:10365–10376

Fourcade Y, Engler JO, Rödder D, Secondi J (2014) Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. PLOS ONE 9:e97122

Frans VF, Augé AA, Fyfe J, Zhang Y and others (2022) Integrated SDM database: enhancing the relevance and utility of species distribution models in conservation management. Methods Ecol Evol 13:243–261

Frisbie MP, Wyman RL (1992) The effect of soil chemistry on sodium balance in the red-backed salamander: a comparison of two forest types. J Herpetol 26:434–442

Government of Canada (2023) Schedule 1 - Species at Risk Act, SC 2002, c 29

Government of Ontario (2023) Species at risk in Ontario list – Endangered Species Act, SO 2007, c 6

Guerry AD, Hunter ML (2002) Amphibian distributions in a landscape of forests and agriculture: an examination of landscape composition and configuration. Conserv Biol 16:745–754

Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I and others (2013) Predicting species distributions for conservation decisions. Ecol Lett 16:1424–1435

Horne MT, Dunson WA (1994a) Behavioral and physiological responses of the terrestrial life stages of the Jefferson salamander, *Ambystoma jeffersonianum*, to low soil pH. Arch Environ Contam Toxicol 27:232–238

Horne MT, Dunson WA (1994b) Exclusion of the Jefferson salamander, *Ambystoma jeffersonianum*, from some potential breeding ponds in Pennsylvania: effects of pH, temperature, and metals on embryonic development. Arch Environ Contam Toxicol 27:323–330

Jaeger RG (1980) Microhabitats of a terrestrial forest salamander. Copeia 1980:265–268

Jarnevich CS, Young N (2015) Using the MAXENT program for species distribution modelling to assess invasion risk. In: Venette RC (ed) Pest risk modelling and mapping for invasive alien species. CABI, Wallingford, p 65–81

Jha A, Praveen J, Nameer PO (2022) Contrasting occupancy models with presence-only models: Does accounting for detection lead to better predictions? Ecol Modell 472: 110105

Jones RP (2022) Ford government forges ahead with Greenbelt development plan despite 'broad opposition' in public consultation. CBC News, 22 Dec 2022. https://www.cbc.ca/news/canada/toronto/greenbelt-oak-ridges-moraine-regulations-1.6692337

Keeley JE, Zedler PH (1998) Characterization and global distribution of vernal pools. In: Witham CW, Bauder ET, Belk D, Ferren WR Jr, Ornduff R (eds) Ecology, conservation, and management of vernal pool ecosystems: proceedings from a 1996 conference. California Native Plant Society, Sacramento, CA, p 1–14

Koma Z, Seijmonsbergen AC, Grootes MW, Nattino F and others (2022) Better together? Assessing different remote sensing products for predicting habitat suitability of wetland birds. Divers Distrib 28:685–699

Lee-Yaw JA, McCune JL, Pironon S, Sheth SN (2022) Species distribution models rarely predict the biology of real populations. Ecography 2022:1–16

Lindsay JB (2016) Whitebox GAT: a case study in geomorphometric analysis. Comput Geosci 95:75–84

Liu C, White M, Newell G (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. J Biogeogr 40:778–789

Lunghi E, Manenti R, Mulargia M, Veith M, Corti C, Ficetola GF (2018) Environmental suitability models predict population density, performance and body condition for microendemic salamanders. Sci Rep 8:7527

McCune JL (2016) Species distribution models predict rare species occurrences despite significant effects of landscape context. J Appl Ecol 53:1871–1879

Merow C, Smith MJ, Silander JA (2013) A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. Ecography 36: 1058–1069

Mills PB, Hossie TJ, Murray DL (2020) Niche determinants in a salamander complex: Does hybridism or reproductive parasitism explain patterns of distribution? Ecosphere 11:e03265

Morales NS, Fernández IC, Baca-González V (2017) MaxEnt's parameter configuration and small samples: Are we paying attention to recommendations? A systematic review. PeerJ5:e3093

Nottingham S, Pelletier TA (2021) The impact of climate change on western *Plethodon* salamanders' distribution. Ecol Evol 11:9370–9384

Olivero J, Toxopeus AG, Skidmore AK, Real R (2016) Testing the efficacy of downscaling in species distribution modelling: a comparison between MaxEnt and favourability function models. Anim Biodivers Conserv 39:99–114

Pelletier TA, Crisafulli C, Wagner S, Zellmer AJ, Carstens BC (2015) Historical species distribution models predict species limits in western *Plethodon* salamanders. Syst Biol 64:909–925

Peterman WE, Semlitsch RD (2013) Fine-scale habitat associations of a terrestrial salamander: the role of environmental gradients and implications for population dynamics. PLOS ONE 8:e62184

Petranka JW (1998) Salamanders of the United States and Canada. Smithsonian Institution Press, Washington, DC

Phillips SJ (2008) Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). Ecography 31:272–278

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Modell 190:231–259

Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME (2017) Opening the black box: an open-source release of Maxent. Ecography 40:887–893

Porej D, Micacchion M, Hetherington TE (2004) Core terrestrial habitat for conservation of local populations of salamanders and wood frogs in agricultural landscapes. Biol Conserv 120:399–409

R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Rains MC, Dahlgren RA, Fogg GE, Harter T, Williamson RJ (2008) Geological control of physical and chemical hydrology in California vernal pools. Wetlands 28:347–362

Rosner-Katz H, McCune JL, Bennett JR (2020) Using stacked SDMs with accuracy and rarity weighting to optimize surveys for rare plant species. Biodivers Conserv 29:3209–3225

Rothermel BB, Semlitsch RD (2006) Consequences of forest fragmentation for juvenile survival in spotted (*Ambystoma maculatum*) and marbled (*Ambystoma opacum*) salamanders. Can J Zool 84:797–807

Srivastava V, Lafond V, Griess VC (2019) Species distribution models (SDM): applications, benefits and challenges in invasive species management. CABI Rev Perspect Agric Vet Sci Nutr Nat Resour 14:1–13

Stevens BS, Conway CJ (2020) Predictive multi-scale occupancy models at range-wide extents: effects of habitat and human disturbance on distributions of wetland birds. Divers Distrib 26:34–48

Sugalski MT, Claussen DL (1997) Preference for soil moisture, soil pH, and light intensity by the salamander, *Plethodon cinereus.* J Herpetol 31:245–250

Taub FB (1961) The distribution of the red-backed salamander, *Plethodon c. cinereus*, within the soil. Ecology 42: 681–698

Trumbo DR, Burgett AA, Hopkins RL, Biro EG, Chase JM, Knouft JH (2012) Integrating local breeding pond, landcover, and climate factors in predicting amphibian distributions. Landsc Ecol 27:1183–1196

van Drunen SG, Linton JE, Bogart JP, McCarter J, Fotherby H, Sandilands A, Norris DR (2020) Estimating critical habitat based on year-round movements of the endangered Jefferson salamander (*Ambystoma jeffersonianum*) and their unisexual dependents. Can J Zool 98:117−126

Warren DL, Seifert SN (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol Appl 21: 335−342

Warren DL, Glor RE, Turelli M (2010) ENMTools: a toolbox for comparative studies of environmental niche models. Ecography 33:607−611

Zhang C, Chen Y, Xu B, Xue Y, Ren Y (2020) Improving prediction of rare species' distribution from community data. Sci Rep 10:12230

Zhou Y, Zhang J (2014) Application of GIS in downscaling regional climate model results over the province of Ontario. Environ Syst Res 3:8