

1 **A new record of *Stylophorum diphyllum* (Michx.) Nutt.) in Canada: a case study of the**
2 **value and limitations of building species distribution models for very rare plants**

3
4 J.L. McCune*

5 University of Guelph

6 Department of Integrative Biology

7 Science Complex

8 50 Stone Road East

9 Guelph, ON N1G 2W1

10

11 *current address:

12 Carleton University

13 Geomatics and Landscape Ecology Research Laboratory

14 Biology Department

15 1125 Colonel By Drive

16 Ottawa, ON K1S 5B6

17 email: jenny.mccune@glel.carleton.ca

18 phone: 519-766-2225

19

20

21

22

23

24 **Abstract.** *Stylophorum diphyllum* (Michx.) Nutt.) is an endangered plant of rich floodplain
25 forests in southern Ontario, Canada. Prior to 2015 there were only four known populations in
26 Ontario. I built a species distribution model (SDM) based on the known occurrences, and tested
27 it by surveying 156 forest sites which varied in their predicted suitability. An indicator species
28 analysis showed that sites predicted to be suitable had significantly higher frequency and
29 abundance of common species usually associated with *S. diphyllum*, demonstrating the ability of
30 the SDM to pinpoint similar habitat, although none of these sites contained *S. diphyllum*. The
31 most important predictors used by the SDM to determine habitat suitability were growing season
32 precipitation, surficial geology, and soil texture. I discovered a new population of *S. diphyllum*
33 more than 50km north of the known populations, at one of the sites not predicted to be suitable.
34 This demonstrates a clear example of SDM overfitting, which may occur when models are built
35 based on few, spatially-limited occurrence records. Nonetheless, the key environmental
36 predictors remained the same in an updated SDM including the new record. *Stylophorum*
37 *diphyllum* provides a case study of both the value and the limitations of using SDMs to predict
38 suitable habitat for very rare and geographically restricted plants, and the need for more rare
39 plant surveys even in human-dominated landscapes.

40

41 Key words: MaxEnt, species distribution model, overfitting, endangered species, *Stylophorum*
42 *diphyllum*.

43

44 **Introduction**

45 *Stylophorum diphyllum* (Michx.) Nutt.) is a perennial wildflower of mature, rich
46 floodplain forests (Jenkins and Parker 2000, Bowles 2007). It is the only species of *Stylophorum*

47 native to North America (COSEWIC 2007). In Canada it was first noted by Robert Elliott in
48 1887, who collected it near the village of Plover Mills, near London, Ontario (Bowles n.d.,
49 McCoun 1888, Grainger 2002). This specimen is deposited in the National Herbarium of Canada
50 (Catalogue CAN 59683). Keddy (1987) thought *S. diphyllum* was likely extinct in Canada
51 because it had not been collected in nearly a century. Sightings were reported (but no collections
52 made) in the early 1970s, and a population was discovered near London, Ontario in 1987
53 (Oldham 1992, COSEWIC 2007). As of 2007 there were three known populations, all within 30
54 km of each other in the London area (COSEWIC 2007).

55 *Stylophorum diphyllum* is quite distinctive in Ontario and elsewhere at the northern edge
56 of its range because it is the only native yellow-flowered poppy in this region (Reznicek *et al.*
57 2011). It flowers in mid- to late-May in Ontario, producing large, bright yellow flowers with four
58 petals. It can grow up to 40 cm in height (Bowles 2007). The stem leaves are opposite and deeply
59 lobed, and the ovary and subsequent fruit are hairy. These hairy ovoid capsules and the opposite
60 leaves distinguish *S. diphyllum* from its exotic relative, Celandine (*Chelidonium majus* L.).

61 *Stylophorum diphyllum* can be locally common towards the centre of its global range,
62 particularly in Kentucky where it is ranked S5 (secure; NatureServe 2015). However, in Canada
63 it is ranked S1 (critically imperilled), and assessed as endangered by the Committee on the Status
64 of Endangered Wildlife in Canada (COSEWIC; COSEWIC 2007). It is listed as Endangered
65 under the *Ontario Endangered Species Act* (Bowles 2011), and the federal *Species at Risk Act*
66 (Bowles 2007). This designation is based on the small number and size of the known
67 populations, decline of habitat due to development and invasive species, and threats of trampling
68 by hikers and all-terrain vehicles (COSEWIC 2007). However, it is not clear why the species is
69 so rare in Canada. It is a polycarpic perennial that can begin producing seed as early as the first

70 year of life, and one plant can produce over 1,000 seeds each year (COSEWIC 2007, Racke
71 2010). It can also successfully colonize areas of disturbed soil – in fact it seems to favour such
72 areas for germination (Bowles 2007). It is available in the nursery trade and does well in gardens
73 (Bowles 2007). Nonetheless, it has not been observed to spread outwards from naturally
74 occupied areas in Ontario, even though the surrounding habitat appears to be suitable
75 (COSEWIC 2007). A transplant experiment in Kentucky found almost no germination of seeds
76 added to suitable habitat (Racke 2010). Seeds are dispersed by ants, who collect the seeds for the
77 oily elaiosome and then discard the rest of the seed (COSEWIC 2007).

78 Potential suitable habitat in the vicinity of the known populations in Ontario has been
79 relatively well searched by botanists, and the 2007 status report notes that: “*The striking*
80 *appearance of this species in early spring makes it unlikely that it has been seriously*
81 *overlooked*” (COSEWIC 2007). Nevertheless, in 2010 an additional population was discovered
82 about 3 km from one of the three known occurrences. It is possible that there are additional
83 undiscovered populations of *S. diphyllum* in southern Ontario.

84 Species distribution models (SDMs) have been used successfully to help direct surveys
85 for new populations of rare species (e.g., Engler *et al.* 2004, Bourg *et al.* 2005, Guisan *et al.*
86 2006, Williams *et al.* 2009, Le Lay *et al.* 2010). In this study, I built an SDM for *S. diphyllum* in
87 southern Ontario based on the four known occurrences, and tested it based on field surveys. I
88 aimed to determine: (1) whether using the SDM predictions could lead to the discovery of new
89 populations of *S. diphyllum*, (2) whether the SDM could successfully direct me to sites with
90 similar habitat conditions to those of the known occurrences, and (3) what environmental
91 conditions differentiate suitable versus unsuitable sites.

92

93 **Materials and Methods**

94 SDMs use the locations of known populations along with spatial data on climate, soils,
95 topography, and other environmental factors to predict other sites in the study area with similar
96 characteristics (Guisan and Zimmerman 2000). A pilot study of eight rare woodland plant
97 species in southern Ontario showed that SDMs were effective for predicting suitable habitat for
98 most species, with success rates of up to 30% (i.e., finding the rare plant at 3 of 10 surveyed sites
99 that were predicted to have suitable habitat; see McCune 2016). In 2015 I built SDMs for an
100 additional set of rare plant species, including *S. diphyllum*.

101 I used the program MaxEnt (version 3.3.3k) to build the SDM for *S. diphyllum*. MaxEnt
102 is a machine-learning based method that has been shown to perform well even with small
103 numbers of known occurrences (Hernandez *et al.* 2006, Phillips *et al.* 2006, Pearson *et al.* 2007,
104 Elith *et al.* 2011). The name MaxEnt refers to the principle of *maximum entropy*, which holds
105 that the best model should be as similar to prior expectations as possible given the available data
106 (Phillips *et al.* 2006, Franklin 2010, Elith *et al.* 2011, Merow *et al.* 2013). In terms of a species'
107 distribution in environmental space, the program estimates the range of environmental conditions
108 the species can tolerate based on the known occurrences, while not narrowing this range more
109 than is warranted by the known occurrences. In other words, "... it is the best approximation of
110 an unknown distribution because it agrees with everything that is known but avoids assuming
111 anything that is unknown" (Franklin 2010, p. 174). MaxEnt is also flexible in that it does not
112 assume that the relationship between a species' occurrence and an environmental predictor is
113 linear, but also considers non-linear relationships when searching for the best model (Elith *et al.*
114 2011). MaxEnt is freely available online
115 (http://biodiversityinformatics.amnh.org/open_source/maxent/).

116 The study extent included all of southern Ontario, and the grain size was one hectare (100
117 x100 m). The choice of study extent is important, and has been shown to affect the accuracy of
118 the resulting SDM (e.g. Franklin 2010, Gogol-Prokurat 2011). Choosing a larger extent, for
119 example all of eastern North America, would have allowed the incorporation of many more
120 occurrence records of *S. diphyllum* from the central portion of its range, and a larger number of
121 occurrence records can increase the quality of SDMs (van Proosdij *et al.* 2016). On the other
122 hand, species with such wide distributions may have different ecological tolerances across their
123 range due to local adaptation of sub-populations, which could make an SDM built across the full
124 geographic extent less accurate when applied to regions at the edge of the range (Stockwell and
125 Peterson 2002). In fact, some studies have found species that are at the edge of their range within
126 the study extent yield more accurate SDMs than species at the centre of their range (e.g. Edwards
127 *et al.* 2005, Luoto *et al.* 2005). The pilot study showed that it was possible to build useful SDMs
128 for plant species at their northern range edge in southern Ontario using a study extent
129 encompassing southern Ontario only, including one species with only 5 occurrence records
130 (McCune 2016), therefore I maintained the same study extent when building the SDM for *S.*
131 *diphyllum*. I used 14 environmental predictors: elevation, slope, aspect, surficial geology type
132 (categorical), soil texture (categorical), soil drainage (categorical), annual mean temperature,
133 mean temperature of the growing season, mean temperature of the wettest quarter, isothermality,
134 annual precipitation, total precipitation for the growing season, precipitation of the warmest
135 quarter, and precipitation seasonality. I obtained climate data from the Canadian Forest Service
136 (McKenney *et al.* 2011). These 14 variables were minimally inter-correlated. For more details,
137 see McCune (2016).

138 I built two SDMs based on the four known locations of *S. diphyllum* in Ontario, which I
139 obtained from the Ontario Natural Heritage Information Centre (NHIC). Each model was
140 identical except I set the regularization multiplier to 1 (the default) for the first model version,
141 and to 0.5 for the second. Regularization is a method used to penalize models that are overly
142 complex, which can lead to models that fit the training data very closely, but are not
143 generalizable (Elith *et al.* 2011, Merow *et al.* 2013). The regularization multiplier sets the
144 strength of regularization, and varying the regularization multiplier is recommended by Merow
145 *et al.* (2013). I used the cumulative model output (Phillips *et al.* 2006, Merow *et al.* 2013) to
146 indicate the predicted suitability of each one-hectare square in southern Ontario. The cumulative
147 value ranges from 0 to 100%, with 100% being the most suitable. I set the threshold for
148 predicting suitable habitat to the minimum output value resulting in the correct prediction of all
149 known locations (i.e., a 0% omission rate) and selected the model version that predicted the
150 lowest percentage of the study area to be suitable (i.e., the “minimal predicted area” or MPA, see
151 Engler *et al.* 2004).

152 During the summers of 2014 and 2015, my field assistants and I surveyed 156 one-
153 hectare forest sites that were predicted to have suitable habitat for one or more of the rare plant
154 species I had modeled, including 13 sites predicted to be suitable for *S. diphyllum*. We used a
155 hand-held GPS unit to navigate to the centre of each one hectare grid cell. We then used a
156 compass and rangefinder to flag straight lines 50m in each cardinal direction from the centre
157 point, delineating four quadrants within each cell. We systematically searched each site by
158 walking straight lines back and forth across each quadrant, maintaining a maximum distance of 5
159 metres between surveyors, until the entire one hectare grid cell had been covered. We noted all
160 vascular plant species present, and estimated the abundance of each according to a coarse

161 abundance scale. If we discovered the target species, we counted the number of individuals,
162 noted whether any were flowering or fruiting, and took photographs. Nomenclature follows
163 Reznicek *et al.* (2011). Survey times ranged from 3 to 10 person hours, depending on the
164 vegetation and the terrain. 110 of the sites were on private land, while 46 were located on land
165 owned by a Conservation Authority, County Forests, or a private Nature Reserve (McCune *et al.*
166 2017).

167 I used the plant community data to determine whether the SDM was predicting the right
168 habitat type, even if *S. diphyllum* was not present. I used an indicator species analysis to
169 determine which species were significantly more frequent or abundant at sites predicted to be
170 suitable for *S. diphyllum* compared to sites not predicted to be suitable (Duf r ne and Legendre
171 1997). I then compared these indicator species to lists of common floodplain forest species and
172 frequent associates of *S. diphyllum*. I used the package ‘labdsv’ in the statistical program R, with
173 5000 randomizations to determine p-values (Roberts 2016, R Core Development Team 2017).

174 MaxEnt provides a ranking of the environmental variables according to their percent
175 contribution to the fit of the model (Phillips 2006). I examined these results and compared the
176 frequency distribution of each predictor variable at all sites predicted to be suitable versus an
177 equal number of randomly-selected one-hectare sites from throughout the study extent. This
178 analysis provided an indication of what environmental conditions differentiate sites predicted to
179 be suitable versus unsuitable for *S. diphyllum*.

180

181 **Results**

182 The SDM model for *S. diphyllum* with the smallest MPA was the model with
183 regularization parameter set to 0.5. The area under the receiver operating curve (AUC) for the

184 test data withheld from the training dataset was 0.978 for this model, which is considered
185 excellent (Swets 1988, Fielding and Bell 1997).

186 The area predicted suitable for *S. diphyllum* comprised about 6,500 hectares (0.05% of
187 the study extent) clustered in the vicinity of the four known records (Fig. 1a). We did not find *S.*
188 *diphyllum* at any of the 13 suitable sites we searched. However, these sites were characterized by
189 indicator species that are typical rich floodplain forest inhabitants, including the natives *Geum*
190 *laciniatum* Murray, *Lysimachia ciliata* L., and *Staphylea trifolia* L., and the exotics *Aegopodium*
191 *podagraria* L. and *Lysimachia nummularia* L.. Indicator species also included seven species
192 noted to be associated with *S. diphyllum* at the known Ontario sites: *Alliaria petiolata* (M. Bieb.)
193 Cavara & Grande, *Celtis occidentalis* L., *Erythronium americanum* Ker Gawl., *Fagus*
194 *grandifolia* Ehrh., *Geranium maculatum* L., *Prunus serotina* Ehrh., and *Ulmus* spp. (Table 1).

195 The environmental variables that contributed the most to the model were total
196 precipitation of the growing season (21.4%), surficial geology (19.1%), soil texture (18.8%),
197 precipitation seasonality (15.9%) and aspect (8.8%). Sites predicted to be suitable for *S.*
198 *diphyllum* receive an average of 659 mm of precipitation in the growing season, while an equal-
199 sized random sample of sites from throughout the study area had an average of 593 mm total
200 precipitation in the growing season (Fig. 2a). Sites predicted to be suitable were more likely to be
201 found on glaciofluvial deposits, modern alluvial deposits, and glaciolacustrine-derived silty to
202 clayey till, and less likely to occur on organic deposits, massive well-laminated glaciolacustrine
203 deposits, or stone-poor, carbonate-derived silty to sandy till. In terms of soil texture, suitable
204 sites were more frequently located on the variable soils of valleys and bottomlands, silt loams,
205 and silty clay loams, and less likely to occur on loam, clay loam, or sandy loam. Suitable sites
206 never occurred on clay or organic soils. Sites predicted to be suitable for *S. diphyllum* had lower

207 average precipitation seasonality than the average across the study area, meaning that they
208 experience less variation in precipitation between months. Finally, sites predicted to be suitable
209 tended to occur on north-facing slopes more than on south-facing slopes.

210 We discovered a new population of *S. diphyllum* at one of the survey sites that was not
211 predicted to be suitable, near the town of Bayfield, more than 50 km north of the known Ontario
212 records (Fig. 1). The site was a mature maple-dominated forest on the Bayfield River floodplain,
213 with components of ash and beech. A number of large old stumps suggested the site had been
214 selectively logged in the past. The *S. diphyllum* population consisted of two clusters about 20 m
215 apart, with a total of approximately 30 multi-stemmed plants. The plants were growing along the
216 edge of an old logging track, on a very slight slope at the base of a steep descent from the fields
217 above down onto the floodplain.

218 We documented a total of 105 vascular plant species within the one-hectare square
219 containing one of the clusters of *S. diphyllum* plants. The understory community was dominated
220 by *Matteuccia struthiopteris* (L.) Todaro and *Solidago flexicaulis* L., with *Geranium maculatum*,
221 *Impatiens capensis* Meerb., *Jeffersonia diphylla* (L.) Pers., *Lithospermum latifolium* Michx.,
222 *Maianthemum stellatum* (L.) Link., and *Ribes cynosbati* L. also common. *Allium tricoccum*
223 Aiton, *Caulophyllum thalictroides* (L.) Michx., *E. americanum*, *J. diphylla*, *M. stellatum*, *R.*
224 *cynosbati*, *S. flexicaulis*, *Trillium* spp., *Tussilago farfara* L., and *Viola* spp. were growing in
225 close proximity to *S. diphyllum* plants. All are native species except for *T. farfara*. We searched
226 two more one-hectare grid squares on the same property, but found no additional *S. diphyllum*
227 plants.

228 After our discovery of the new population, I built an updated SDM by including the new
229 location. Just adding this small amount of information substantially increased the area predicted

230 to be suitable (Fig. 1a,b). Nevertheless, the most important variables for distinguishing suitable
231 habitat from non-suitable habitat were largely the same as for the first model, with suitable sites
232 tending to occur in areas with higher total precipitation in the growing season (Fig. 2b), lower
233 precipitation seasonality, and on the geology and soil types listed above. Only aspect was no
234 longer an important predictor.

235

236 **Discussion**

237 The first step in conserving threatened plant species is to understand where their
238 populations are, their status, and the threats they are facing. Although southern Ontario is highly
239 populated, with forests extremely reduced and fragmented, there are still gaps in this basic
240 knowledge for many threatened plants. Targeting suitable habitat to survey is one way to fill this
241 gap, and this can be done using expert knowledge or by using SDMs to predict locations of
242 potentially suitable habitat.

243 Many studies using SDMs to target locations to survey for rare plant species have led to
244 the discovery of new populations in areas predicted to be suitable (e.g. Van Manen *et al.* 2002,
245 Boetsch *et al.* 2003, Engler *et al.* 2004, Bourg *et al.* 2005, Van Manen *et al.* 2005, Guisan *et al.*
246 2006, Aitken *et al.* 2007, Williams *et al.* 2009, Le Lay *et al.* 2010, Sarkinen *et al.* 2013, McCune
247 2016). Those that have compared SDM-based sampling to random sampling strategies have
248 found SDM-based sampling increases the odds of finding the target species up to 12 times (Van
249 Manen *et al.* 2005, Guisan *et al.* 2006). While expert opinion can also be used to direct surveys
250 to suitable habitat, SDMs can target rare plant surveys when such expert opinion is lacking, or
251 supplement it in cases where it may be biased (Van Manen *et al.* 2002).

252 The MaxEnt model for *S. diphyllum* did not result in the discovery of any new
253 populations. However, this does not in itself indicate a useless SDM. Many studies that have
254 used SDMs to target suitable habitat for rare plants have found that sites with apparently suitable
255 habitat often lack the species (e.g. Van Manen *et al.* 2002, Williams *et al.* 2009, McCune 2016).
256 Even if the habitat is truly suitable, the species might not occupy a site: for example, due to
257 dispersal limitation or interactions with other species (e.g. competitors or parasites; Boetsch *et al.*
258 2003, Gogol-Prokurat 2011). This is why it is important – particularly for rare species - to have
259 other means of measuring habitat suitability in addition to the presence or absence of the target
260 species. Plant community data suggests that the SDM for *S. diphyllum* successfully predicted the
261 right sort of habitat, with the 13 suitable sites characterized by a significantly higher frequency
262 and/or abundance of many common species reported to be associated with *S. diphyllum*.
263 Therefore, although the predictions of the SDM did not lead to the discovery of new populations
264 of *S. diphyllum*, it effectively predicted the right sort of habitat in the vicinity of the known
265 occurrences.

266 When the known occurrences of a species are very few, and particularly when they are
267 spatially clustered, SDM overfitting may occur (Vaughan and Ormerod 2005, Newbold *et al.*
268 2010, Merow *et al.* 2013, Breiner *et al.* 2015). Overfitting results in a model that fits the training
269 data very well, but performs poorly when predicting new locations outside the region of the
270 known occurrences (Vaughan and Ormerod 2005, Peterson *et al.* 2007, Breiner *et al.* 2015). The
271 large increase in the extent of habitat predicted to be suitable by the SDM that included the
272 newly discovered outlier *S. diphyllum* population in the training data indicates that the original
273 SDM was overfit (Fig. 1a,b). Nevertheless, the most important variables for distinguishing
274 suitable habitat from non-suitable habitat were largely the same as for the first model, with

275 suitable sites tending to occur in areas with higher total precipitation in the growing season (Fig.
276 2b), lower precipitation seasonality, and on the geology and soil types listed above. Only aspect
277 was no longer an important predictor. An SDM built with a very low number of occurrences
278 should not be interpreted as representing the actual range limits of a species, but only as an
279 indication of areas where habitat conditions are similar to those at known locations (Pearson *et*
280 *al.* 2007). Therefore, although field surveys based on SDM predictions for very rare plant species
281 may not result in a high discovery rate, the SDM can direct searches to the right habitat type in
282 the vicinity of known occurrences, and help to determine which environmental factors are
283 important for the species' occurrence.

284 To be most effective, predicting suitable habitat with SDMs must be an iterative process
285 (Guisan *et al.* 2006). The potential for model overfitting is high for species with few occurrence
286 records available. There is some evidence that MaxEnt is particularly prone to overfitting
287 (Peterson *et al.* 2007), even though the default settings of MaxEnt and the use of regularization
288 have been designed to reduce this problem (Phillips *et al.* 2006, Merow *et al.* 2013). Future
289 attempts to improve the SDM for *S. diphyllum* could try increasing the regularization multiplier
290 (which increases the penalty against overly-complex models, see Merow *et al.* 2013), or perhaps
291 use the “ensemble of small models” model-averaging technique (Breiner *et al.* 2015). Some
292 authors have suggested filtering occurrence records such that they are a minimum distance apart
293 and not spatially autocorrelated to reduce overfitting (e.g. Shcheglovitova and Anderson 2013).
294 However, this approach is not an option with only 5 available occurrence records. By including
295 the new population of *S. diphyllum* into a revised SDM, I was able to show a large increase in the
296 predicted potential range of habitat suitability for this species. Further field surveys and
297 additional new populations will allow this new SDM to be tested and improved.

298 The newly discovered population does not expand the known latitudinal range of *S.*
299 *diphyllum*: the species is found on the northwest corner of Michigan's Lower Peninsula,
300 approximately 190km north in latitude from the Bayfield population (COSEWIC 2007, Reznicek
301 *et al.* 2011). However, it is a significant expansion of the Canadian range, as all known and
302 historical locations are in a very restricted area of the Thames River watershed; it was considered
303 unlikely that the species ever existed elsewhere in the province (COSEWIC 2007). It is possible
304 that the new population is introduced from nursery stock. However, the site is approximately
305 650m from the nearest house or garden in a mature forest with relatively low cover of introduced
306 species. Research into the genetic variability of all Ontario wood-poppy populations and their
307 degree of relatedness to populations in the U.S. and nursery stock has not been completed
308 (Bowles 2007, S. Snyder, Ontario NHIC, *personal communication* 2017). Without genetic
309 analysis of the new population it is impossible to know its relationship to other wild or
310 introduced populations.

311 The discovery of a new population of *S. diphyllum* shows that undiscovered populations
312 of threatened plant species still exist, even on landscapes that are human-dominated and
313 considered to be relatively well-botanized. For example, the Atlas of Rare Vascular Plants of
314 Ontario (Argus *et al.* 1982-1987) includes 29 taxa that were known only from collections prior to
315 1925, all of which were presumed to be extirpated from the province. Since the book was
316 published, nine of these (31%) have been rediscovered (Argus 1992). Although the situation for
317 threatened plants in human-dominated landscapes is precarious, they can still surprise us. SDMs
318 for extremely rare species with few, spatial clustered populations are prone to overfitting, and
319 may therefore under-predict the true extent of suitable habitat. Nonetheless, they can help

320 pinpoint sites with potentially suitable habitat in the vicinity of known populations, and suggest
321 which environmental predictors might be important in determining habitat suitability.

322

323 **Acknowledgements**

324 Thanks to my excellent field crew: Amber Lammers, Kathryn Tisshaw, Alyson Van Natto, and
325 Melissa Strub. Michael Oldham, Andrew MacDougall, and Anne Bell were my advisors during
326 this project. Michael Oldham and Carole Ann Lacroix helped with species' identifications.

327 Permission to access land was granted by: Ausable Bayfield Conservation Authority, Bruce Trail
328 Association, Conservation Halton, Grand River Conservation Authority, Hamilton Conservation
329 Authority, Hamilton Naturalists Club, Lanark County Forests, Long Point Basin Land Trust,
330 Ontario Heritage Foundation, Niagara Peninsula Conservation Authority, Upper Thames River
331 Conservation Authority, and 102 private landowners. In particular I would like to thank G.C.
332 Ebers, Jennifer Pate, and Sharon Vitali for their support. Funding was provided by the Liber Ero
333 Fellowship Program and the Ontario Species at Risk Stewardship Fund. This project also
334 received financial support from the Government of Canada through the federal Department of
335 the Environment. The views expressed in this paper are the author's and do not necessarily
336 represent those of the funding bodies.

337

338 **Literature Cited**

339 AITKEN, M., D. W. ROBERTS, AND L. M. SHULTZ. 2007. Modeling distributions of rare plants in
340 the Great Basin, western North America. *West. N. Am. Naturalist* 67: 26-38.

341 ARGUS, G.W. 1992. The phytogeography of rare vascular plants in Ontario and its bearing on
342 plant conservation. *Can. J. Bot.* 70: 469-490.

- 343 ARGUS, G.W., K.M. PRYER, D.J. WHITE, AND C.J. KEDDY. eds. 1982-1987. Atlas of the Rare
344 Vascular Plants of Ontario. Parts I, II, III and IV. National Museum of Natural Sciences,
345 Ottawa.
- 346 BOETSCH, J. R., F. K. VAN MANEN, AND J. D. CLARK. 2003. Predicting rare plant occurrence in
347 Great Smoky Mountains National Park, USA. *Nat. Area. J.* 23:229-237.
- 348 BOURG, N.A., W.J. MCSHEA, AND D.E. GILL. 2005. Putting a cart before the search: successful
349 habitat prediction for a rare forest herb. *Ecology* 86: 2793-2804.
- 350 BOWLES, J. no date. *Stylophorum diphyllum*, wood poppy – rare for a reason – 4(3). Retrieved
351 July 2016 from The North American Native Plant Society Website.
352 <[http://www.nanps.org/index.php/gardening/native-plants-to-know/117-stylophorum-](http://www.nanps.org/index.php/gardening/native-plants-to-know/117-stylophorum-diphyllum-wood-poppy-rare-for-a-reason)
353 [diphyllum-wood-poppy-rare-for-a-reason](http://www.nanps.org/index.php/gardening/native-plants-to-know/117-stylophorum-diphyllum-wood-poppy-rare-for-a-reason)>.
- 354 BOWLES, J.M. 2007. Recovery Strategy for Wood-poppy (*Stylophorum diphyllum*) in Canada.
355 Retrieved July 2016 from The Species at Risk Public Registry Website.
356 <[http://www.registrelep-sararegistry.gc.ca/virtual_sara/files/plans/rs_wood-](http://www.registrelep-sararegistry.gc.ca/virtual_sara/files/plans/rs_wood-poppy_final_0507_e.pdf)
357 [poppy_final_0507_e.pdf](http://www.registrelep-sararegistry.gc.ca/virtual_sara/files/plans/rs_wood-poppy_final_0507_e.pdf)>.
- 358 BOWLES, J.M. 2011. Recovery strategy for the wood-poppy (*Stylophorum diphyllum*) in Ontario.
359 Ontario Recovery Strategy Series. Prepared for the Ontario Ministry of Natural
360 Resources, Peterborough, Ontario.
- 361 BREINER, F.T., A. GUISAN, A. BERGAMINI, and M.P. NOBIS. 2015. Overcoming limitations of
362 modelling rare species by using ensembles of small models. *Methods Ecol. Evol.* 6:
363 1210-1218.

- 364 COSEWIC. 2007. COSEWIC assessment and update status report on the wood-poppy
365 *Stylophorum diphyllum* in Canada. Retrieved July 2016 from The Species at Risk Public
366 Registry Website. <www.sararegistry.gc.ca/status/status_e.cfm>.
- 367 DUFRÈNE, M., AND P. LEGENDRE. 1997. Species assemblages and indicator species: the need for
368 a flexible asymmetrical approach. *Ecol. Monogr.* 67: 345–366.
- 369 EDWARDS, T. C., D. R. CUTLER, N. E. ZIMMERMAN, L. GEISER, AND J. ALEGRIA. 2005. Model-
370 based stratifications for enhancing the detection of rare ecological events. *Ecology*
371 86:1081-1090.
- 372 ELITH, J., S.J. PHILLIPS, T. HASTIE, M. DUDIK, Y.E. CHEE, AND C.J. YATES. 2011. A statistical
373 explanation of MaxEnt for ecologists. *Divers. Distrib.* 17: 43-57.
- 374 ENGLER, R., A. GUISAN, AND R. RECHSTEINER. 2004. An improved approach for predicting the
375 distribution of rare and endangered species from occurrence and pseudo-absence data. *J.*
376 *Appl. Ecol.* 41: 263-274.
- 377 FIELDING, A. H., AND J.F. BELL. 1997. A review of methods for the assessment of prediction
378 errors in conservation presence/absence models. *Environ. Conserv.* 24: 38-49.
- 379 FRANKLIN, J. 2010. *Mapping Species Distributions: spatial inference and prediction*. Cambridge
380 University Press, Cambridge.
- 381 GOGOL-PROKURAT, M. 2011. Predicting habitat suitability for rare plants at local spatial scales
382 using a species distribution model. *Ecol. Appl.* 21:33-47.
- 383 GRAINGER, J. 2002. *Vanished villages of Middlesex*. Natural Heritage/Natural History Inc.,
384 Toronto, ON. 383 p.

- 385 GUISAN, A., O. BROENNIMANN, R. ENGLER, M. VUST, N.G. YOCCOZ, A. LEHMANN, AND N.E.
386 ZIMMERMAN. 2006. Using niche-based models to improve the sampling of rare species.
387 Conserv. Biol. 20: 501-511.
- 388 GUISAN, A., AND N.E. ZIMMERMANN. 2000. Predictive habitat distribution models in ecology.
389 Ecol. Model. 135: 147-186.
- 390 HERNANDEZ, P.A., C.H. GRAHAM, L.L. MASTER, AND D.L. ALBERT. 2006. The effect of sample
391 size and species characteristics on performance of different species distribution modeling
392 methods. Ecography 29: 773-785.
- 393 JENKINS, M.A., AND G.R. PARKER. 2000. The response of herbaceous-layer vegetation to
394 anthropogenic disturbance in intermittent stream bottomland forests of southern Indiana,
395 USA. Plant Ecol. 151(2): 223-337.
- 396 KEDDY, C.J. 1987. *Stylophorum diphyllum* (Michx.) Nutt. In Atlas of the Rare Vascular Plants of
397 Ontario. Edited by G. W. Argus, K. M. Pryer, D. J. White and C. J. Keddy. Part IV.
398 National Museum of Natural Sciences, Ottawa.
- 399 LE LAY, G., R. ENGLER, E. FRANC, AND A. GUISAN. 2010. Prospective sampling based on model
400 ensembles improves the detection of rare species. Ecography 33: 1015-1027.
- 401 LUOTO, M., J. POYRY, R. K. HEIKKINEN, AND K. SAARINEN. 2005. Uncertainty of bioclimate
402 envelope models based on the geographical distribution of species. Global Ecol.
403 Biogeogr. 14:575-584.
- 404 MCCOUN, J. 1888. Catalogue of Canadian Plants. Part IV. – Endogens. Dawson Brothers,
405 Montreal.
- 406 MCCUNE, J.L. 2016. Species distribution models predict rare species occurrences despite
407 significant effects of landscape context. J. Appl. Ecol. 53: 1871-1879.

- 408 MCCUNE, J.L., A. VAN NATTO, AND A.S. MACDOUGALL. 2017. The efficacy of protected areas
409 and private land for plant conservation in a fragmented landscape. *Landscape Ecol.* 32:
410 871-882.
- 411 MCKENNEY, D.W., M.R. HUTCHINSON, P. PAPADOPOL, K. LAWRENCE, J. PEDLAR, K.
412 CAMPBELL, E. MILEWSKA, R.F. HOPKINSON, D. PRICE, AND T. OWEN. 2011. Customized
413 spatial climate models for North America. *B. Am. Meteorol. Soc.*, December: 1611-1622.
- 414 MEROW, C., M.J. SMITH, AND J.A. SILANDER. 2013. A practical guide to MaxEnt for modeling
415 species' distributions: what it does, and why inputs and settings matter. *Ecography* 36:
416 1058-1069.
- 417 NATURESERVE. 2015. NatureServe Explorer: an online encyclopedia of life. Version 7.1.
418 Retrieved July 2017 from the NatureServe Website. <<http://explorer.natureserve.org>>.
- 419 NEWBOLD, T., T. READER, A. EL-GABBAS, W. BERG, W. M. SHOHDHI, S. ZALAT, S. B. EL DIN, AND
420 F. GILBERT. 2010. Testing the accuracy of species distribution models using species
421 records from a new field survey. *Oikos* 119:1326-1334.
- 422 OLDHAM, M.J. 1992. Some recent Middlesex County plant discoveries. *The Cardinal* 146: 24-27.
- 423 PEARSON, R.G., C.J. RAXWORTHY, M. NAKAMURA, and A.T. PETERSON. 2007. Predicting
424 species distributions from small numbers of occurrence records: a test case using cryptic
425 geckos in Madagascar. *J. Biogeogr.* 34: 102-117.
- 426 PETERSON, A.T., M. PAPEŞ, AND M. EATON, M. 2007. Transferability and model evaluation in
427 ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, 30: 550-560.
- 428 PHILLIPS, S. 2006. A brief tutorial on Maxent. AT&T Research. Retrieved July 2016 from
429 Princeton University Website.
430 <<http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc>>.

- 431 PHILLIPS, S.J., R.P. ANDERSON, AND R.E. SCHAPIRE. 2006. Maximum entropy modeling of
432 species geographic distributions. *Ecol. Model.* 190: 231-259.
- 433 RACKE, D. 2010. Reestablishing diversity in our hardwood forests: a transplant study of five
434 spring-flowering herbs. M.Sc. thesis. Western Kentucky University, Bowling Green,
435 Kentucky, USA.
- 436 R DEVELOPMENT CORE TEAM. 2016. R: A Language and Environment for Statistical Computing.
437 Retrieved October 2016 from R Foundation for Statistical Computing, Vienna, Austria.
438 <<https://www.R-project.org/>>.
- 439 REZNICEK, A.A, E.G. VOSS, AND B.S. WALTERS. 2011. Michigan Flora Online. Retrieved March
440 2017 from University of Michigan.
441 <<http://www.michiganflora.net/species.aspx?id=1902>>.
- 442 ROBERTS, D.W. 2012. Package “labdsv.” Retrieved July 2016 from R Core Development team.
443 <<https://cran.r-project.org/web/packages/labdsv/labdsv.pdf>>.
- 444 SARKINEN, T., P. GONZALES, AND S. KNAPP. 2013. Distribution models and species discovery:
445 the story of a new *Solanum* species from the Peruvian Andes. *PhytoKeys* 31:1-20.
- 446 SHCHEGLOVITOVA, M., AND R.P. ANDERSON. 2013. Estimating optimal complexity for ecological
447 niche models: a jackknife approach for species with small sample sizes. *Ecol. Model.*
448 269: 9-17.
- 449 STOCKWELL, D. R. B. AND A. T. PETERSON. 2002. Effects of sample size on accuracy of species
450 distribution models. *Ecol. Model.* 148:1-13.
- 451 SWETS, J.A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240: 1285–1293.
- 452 VAN MANEN, F. T., J. D. CLARK, S. E. SCHLARBAUM, K. JOHNSON, AND G. TAYLOR. 2002. A
453 model to predict the occurrence of surviving butternut trees in the southern Blue Ridge

- 454 mountains. Pages 491-497 in J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler,
455 M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting Species Occurrences:
456 Issues of Accuracy and Scale. Island Press, Washington, DC.
- 457 VAN MANEN, F. T., J. A. YOUNG, C. A. THATCHER, W. B. CASS, AND C. ULREY. 2005. Habitat
458 models to assist plant protection efforts in Shenandoah National Park, Virginia, USA. Nat.
459 Area. J. 25:339-350.
- 460 VAN PROOSDIJ, A. S., M.S. SOSEF, J.J. WIERINGA, AND N. RAES. 2016. Minimum required number
461 of specimen records to develop accurate species distribution models. Ecography 39: 542-
462 552.
- 463 VAUGHAN, I. P. AND S. J. ORMEROD. 2005. The continuing challenges of testing species
464 distribution models. J. Appl. Ecol. 42:720-730.
- 465 WILLIAMS, J.N., C. SEO, J. THORNE, J.K. NELSON, S. ERWIN, J.M. O'BRIEN, and M.W.
466 SCHWARTZ. 2009. Using species distribution models to predict new occurrences for rare
467 plants. Divers. Distrib. 15: 565-576.

468 **Tables**

469 Table 1. Significant indicator species according to an indicator species analysis comparing sites predicted to be suitable
 470 *Stylophorum diphyllum* (N=13) to sites not predicted to be suitable (N=143). Significance was based on 5,000 permutations.
 471 Abundance was ranked on a categorical scale from 1 (very rare), to 5 (dominant).

| Species | Frequency in suitable plots (%) | Frequency in unsuitable plots (%) | Average abundance in suitable plots | Average abundance in unsuitable plots | Indicator of which site |
|-------------------------------|---------------------------------------|---|---|---|----------------------------|
| <i>Actaea pachypoda</i> | 8 | 53 | 0.77 | 1.12 | unsuitable |
| <i>Acer negundo</i> | 85 | 40 | 0.67 | 0.99 | suitable |
| <i>Aegopodium podagraria</i> | 15 | 2 | 0.90 | 0.05 | suitable |
| <i>Allium canadense</i> | 31 | 7 | 0.46 | 0.16 | suitable |
| <i>Alliaria petiolata</i> | 100 | 75 | 0.59 | 2.33 | suitable |
| <i>Allium tricoccum</i> | 46 | 19 | 0.75 | 0.39 | suitable |
| <i>Angelica atropurpureum</i> | 46 | 12 | 0.77 | 0.25 | suitable |
| <i>Barbarea vulgaris</i> | 38 | 9 | 0.69 | 0.16 | suitable |
| <i>Caltha palustris</i> | 38 | 11 | 1.08 | 0.28 | suitable |

| | | | | | |
|----------------------------------|----|----|------|------|----------|
| <i>Cardamine diphyllum</i> | 46 | 8 | 0.77 | 0.19 | suitable |
| <i>Cardamine pensylvanicum</i> | 15 | 1 | 0.31 | 0.02 | suitable |
| <i>Celtis occidentalis</i> | 54 | 14 | 1.00 | 0.29 | suitable |
| <i>Claytonia virginiana</i> | 23 | 3 | 0.38 | 0.08 | suitable |
| <i>Crataegus</i> spp. | 77 | 50 | 2.38 | 1.21 | suitable |
| <i>Equisetum</i> spp. | 85 | 56 | 2.08 | 1.41 | suitable |
| <i>Erythronium americanum</i> | 62 | 12 | 2.08 | 0.36 | suitable |
| <i>Euonymus obovatus</i> | 77 | 38 | 2.00 | 0.90 | suitable |
| <i>Fagus grandifolia</i> | 92 | 64 | 3.08 | 1.82 | suitable |
| <i>Geranium maculatum</i> | 77 | 42 | 2.31 | 1.17 | suitable |
| <i>Geum laciniatum</i> | 85 | 33 | 2.31 | 0.92 | suitable |
| <i>Glechoma hederacea</i> | 38 | 6 | 0.77 | 0.13 | suitable |
| <i>Hesperis matronalis</i> | 85 | 47 | 2.46 | 1.27 | suitable |
| <i>Lysimachia ciliata</i> | 69 | 39 | 1.46 | 0.83 | suitable |
| <i>Lysimachia nummularia</i> | 77 | 40 | 2.46 | 1.20 | suitable |
| <i>Narcissus pseudonarcissus</i> | 15 | 1 | 0.23 | 0.01 | suitable |

| | | | | | |
|------------------------------|-----|----|------|------|----------|
| <i>Podophyllum peltatum</i> | 85 | 51 | 2.46 | 1.43 | suitable |
| <i>Populus tremuloides</i> | 54 | 20 | 1.00 | 0.45 | suitable |
| <i>Prunus serotina</i> | 92 | 69 | 2.38 | 1.64 | suitable |
| <i>Rhamnus cathartica</i> | 100 | 64 | 2.69 | 1.61 | suitable |
| <i>Ribes americana</i> | 46 | 19 | 1.31 | 0.38 | suitable |
| <i>Rumex</i> spp. | 62 | 24 | 1.15 | 0.46 | suitable |
| <i>Staphylea triphylla</i> | 15 | 5 | 0.46 | 0.09 | suitable |
| <i>Symplocarpus foetidus</i> | 69 | 31 | 2.46 | 1.01 | suitable |
| <i>Thalictrum dioicum</i> | 69 | 41 | 1.62 | 1.01 | suitable |
| <i>Trillium</i> spp. | 31 | 8 | 0.62 | 0.15 | suitable |
| <i>Ulmus</i> spp. | 92 | 67 | 2.08 | 1.58 | suitable |
| <i>Urtica dioica</i> | 77 | 40 | 1.77 | 0.98 | suitable |
| <i>Uvularia sessilifolia</i> | 23 | 3 | 0.54 | 0.08 | suitable |
| <i>Veronica persica</i> | 15 | 2 | 0.38 | 0.04 | suitable |

472

473

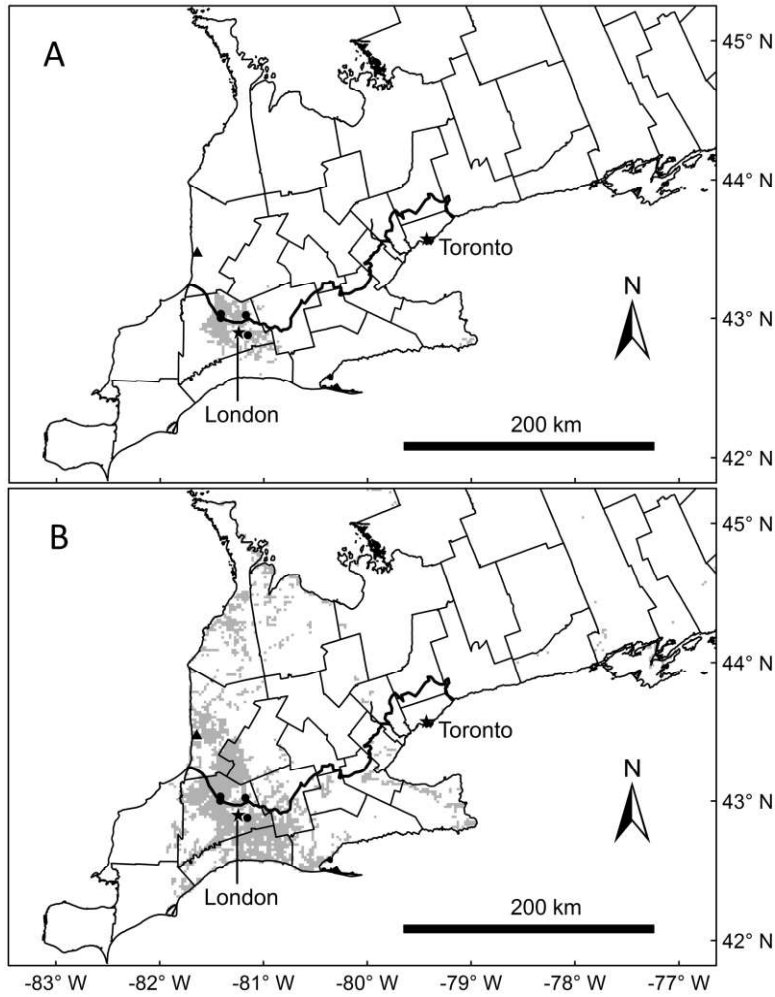
474

475 **Figure Captions**

476 FIGURE 1. The four populations of *Stylophorum diphyllum* in Ontario known prior to 2015 (black circles) and t
477 found in 2015 (black triangle). Thin lines indicate county borders, the thick black line shows the approximate no
478 Carolinian forest zone. Grey shading shows areas predicted by the SDM to have suitable habitat for *S. diphyllum*
479 model was based only on the 4 pre-2015 records, in panel (b) the model included the new 2015 record. Suitable
480 generalized to 2km x 2km squares to increase visibility.

481
482 FIGURE 2. Kernel density plots showing the distribution of total growing season precipitation among one hectare
483 be suitable for *Stylophorum diphyllum* compared to the distribution of an equal sized random sample of squares n
484 suitable. In panel (a) the species distribution model was based only on the 4 pre-2015 records, in panel (b) the m
485 2015 record.

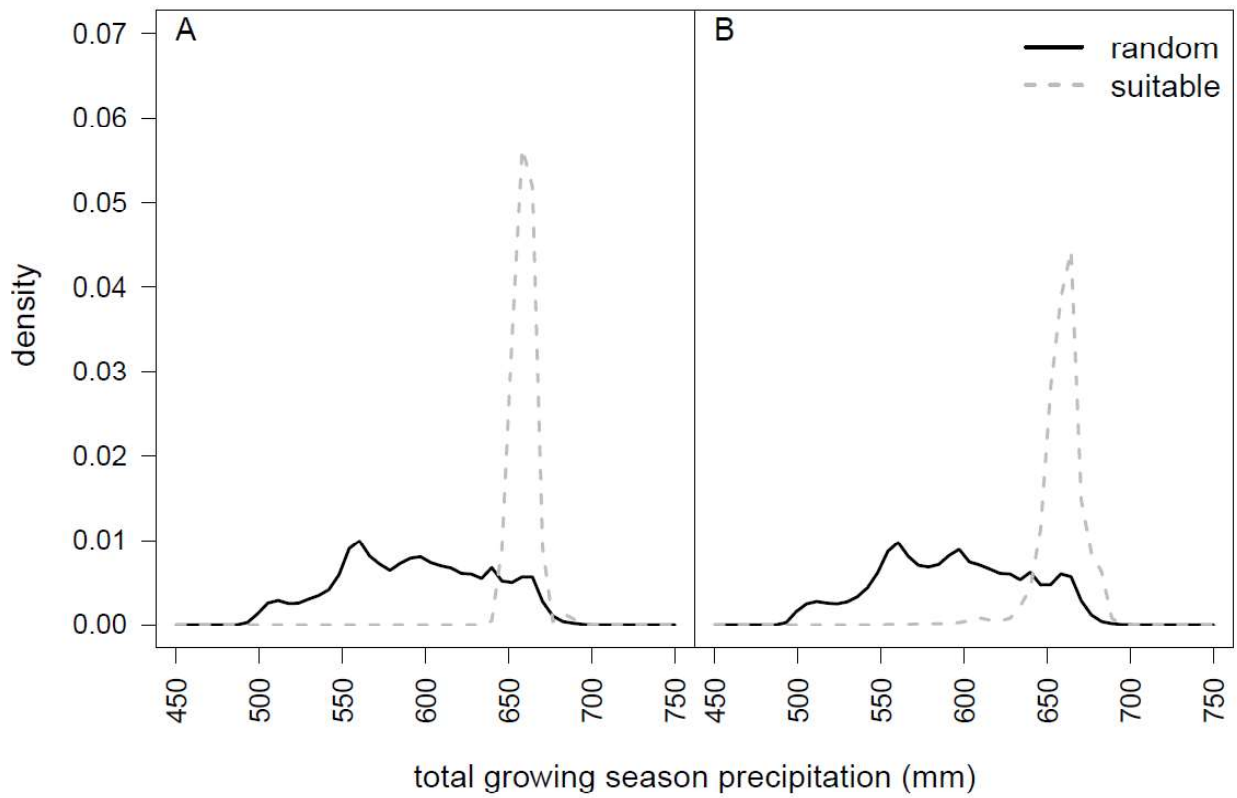
486
487
488
489
490
491



492

493 Figure 1.

494



495

496

497 Figure 2.

498