

1 Noninvasive low-cost method to identify armadillos' burrows: 2 A machine learning approach

3 THIAGO F. RODRIGUES,^{1,2} *Applied Ecology Program, Luiz de Queiroz College of Agriculture, University*
4 *of São Paulo, Av. Pádua Dias 11, Piracicaba, SP 13418-900, Brazil*

5 KEILLER NOGUEIRA, *Data Science Research Group, Computing Science and Mathematics Division,*
6 *University of Stirling, Scotland, FK9 4LA, UK*

7 ADRIANO G. CHIARELLO, *Department of Biology, Faculty of Philosophy, Sciences and Languages of*
8 *Ribeirão Preto, University of São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, SP 14040-901, Brazil*

9
10 **ABSTRACT** Having accurate information about population parameters of armadillos (Mammalia,
11 Cingulata) is essential for the conservation and management of this taxon, most species of which
12 remain poorly studied. We investigated whether we could accurately identify 4 armadillo species
13 (*Euphractus sexcinctus*, *Dasyops novemcinctus*, *Cabassous tatouay*, and *Cabassous unicinctus*) based on their
14 burrow morphometry. We first selected published studies that reported measurements of width,
15 height, and angle of the burrows used by the 4 species of armadillos. Then, using such data we simulate
16 burrow measurements for each of the 4 species of armadillos, we created predictive models through
17 supervised machine learning that were capable of correctly identifying the species of armadillos based
18 on their burrows' morphometry. By using classification algorithms such as Random Forest, K-
19 Nearest Neighbor, Support Vector Machine, Naive Bayes, and Decision Tree C5.0, we achieved the
20 overall accuracy for the classification task by about 71%, including an overall Kappa index by about
21 61%. *Euphractus sexcinctus* was the most difficult species to discriminate and classify (approximately
22 68% of accuracy), whereas *C. unicinctus* was the easiest to discriminate (approximately 93% of
23 accuracy). We found that it was possible to identify similar-sized armadillos based on the

24 measurements of their burrows described in the literature. Finally, we developed an R function
25 (armadilloID) that automatically identified the 4 species of armadillos using burrow morphology. As
26 the data we used represented all studies that reported the morphometry of burrows for the 4 species
27 of armadillos, we can generalize that our function can predict armadillo species beyond our data.

28

29 **KEY WORDS** burrow, conservation, *Cabassous unicinctus*, *Cabassous tatouay*, *Dasyops novemcinctus*,
30 *Euphractus sexcinctus*, mammal, noninvasive method, Xenarthra

31

32 Estimating population parameters for wildlife is one of the primary interests of scientists and
33 conservationists because of its decisive influence on wildlife management and conservation.

34 However, estimating such parameters (e.g., density, abundance, occurrence) is not straightforward,
35 particularly for wide-ranging, low-density, elusive, and unstudied species, most of which are
36 threatened (Schipper et al. 2008, Desbiez et al. 2018). For example, according to the IUCN Red List
37 of Threatened Species, 10 out of the 20 extant armadillo species had unknown population trend
38 information, whilst the other 6 were in decline; 5 species were categorized as Data Deficient, 5 as
39 Near Threatened, and 2 as Vulnerable (IUCN 2019). Because it is not practical to employ tagging
40 methods to derive armadillo population estimates due to logistical difficulties, high costs, and small
41 numbers of possible captures (Loughry and McDonough 2013, Desbiez et al. 2018), less than 20%
42 of armadillo studies were based on fieldwork conducted on wild populations (Superina et al. 2014).
43 Alternatively, noninvasive methods such as camera trapping have been successfully used to assess
44 occupancy for armadillo species (Zimbres et al. 2013, Rodrigues and Chiarello 2018). However,
45 camera-survey methods require a large number of cameras and high associated costs (e.g., Rodrigues
46 and Chiarello 2018), cameras are vulnerable to animal damage, adverse weather, and theft or

47 vandalism, and such methods are likely produce a small number of detections, particularly for rarer
48 species (Maccarini et al. 2015, Desbiez et al. 2018).

49 Although armadillos are observed in nature above ground, they dig burrows for shelter
50 (housing, raising offspring), protection (to hide from predators and to buffer against environmental
51 temperatures), and feeding (foraging burrows) (McNab 1980, Eisenberg and Redford 1999, Desbiez
52 et al. 2018). Burrows created by armadillos have specific shapes and sizes (Carter and Encarnaçao
53 1983, Abba et al. 2005, Trovati 2015, Attias et al. 2016, Desbiez et al. 2018), which are typically
54 influenced by anatomical and morphological differences among species (e.g., Carter and Encarnaçao
55 1983, Attias et al. 2016). Several previous studies have used armadillo burrows to estimate population
56 parameters including habitat use and density and to assess activity and behavior (Zimmerman 1990,
57 McDonough et al. 2000, Abba et al. 2005, 2007, Desbiez et al. 2018). However, identifying
58 armadillos based on their burrows is challenging (McDonough et al. 2000, Arteaga and Venticinque
59 2010). For instance, Carter and Encarnaçao (1983) fitted 4 species of armadillos (*Cabassous*
60 *tatouay*, *Cabassous unicinctus*, *Euphractus sexcinctus*, and *Priodontes maximus*) with radio transmitters and
61 found that the shape of the burrow entrances differed between the species. Not surprisingly, more
62 records on burrow measurements are available for the species with the widest geographic range, the
63 wide-ranging *Dasybus novemcinctus* (Zimmerman 1990, McDonough et al. 2000, Platt et al. 2004,
64 Sawyer et al. 2012). In addition to *D. novemcinctus*, several additional species of armadillos were also
65 monitored and had measurements of their burrows reported (Medri 2008, Attias et al. 2016, Desbiez
66 et al. 2018). However, the relationship between the reported burrow measurements and species
67 identification remains **tenuous** (e.g., Arteaga and Venticinque 2010). Few studies reported burrow
68 measurements for different armadillo species (e.g., Carter and Encarnaçao 1983), and each reported
69 measure can be different among the studies or is not always associated with their corresponding
70 estimate of precision (variance or standard deviation), further complicating the identification process

71 (Carter and Encarnação 1983, McDonough et al. 2000, Medri 2008, Arteaga and Venticinque 2010,
72 Trovati 2015).

73 Machine learning (ML) enables computers to solve tasks analyzing complex patterns without
74 being explicitly programmed to solve those tasks (Sen 2018). State-of-the-art methods teach
75 machines via supervised learning (i.e., by showing them correct pairs of inputs and outputs from
76 labeled data), unsupervised learning (i.e., finding hidden information or structure from unlabeled
77 data), and semi-supervised learning (i.e., a combination of supervised and unsupervised ML
78 technique) (Sen 2018). Machine learning algorithms and models aim to maximize predictability based
79 on data and have demonstrated high accuracy in predicting ecological patterns (Olden et al. 2008,
80 Crisci et al. 2012, Thessen 2016). Machine learning models have been applied increasingly in ecology,
81 including in studies of species distribution modeling (Elith et al. 2006, Phillips et al. 2006), species
82 diversity (Olden et al. 2008), and distribution (Elith and Leathwick 2009), and represent a potential
83 for improving species identification methods (Norouzzadeh et al. 2018).

84 Considering that *C. tatouay*, *C. unicinctus*, *D. novemcinctus*, and *E. sexcinctus* have overlapping
85 ranges and inhabit regions affected by anthropogenic disturbances (Vivo et al. 2011, Egeskog et al.
86 2014, Trovati 2015), improving identification methods might help answer questions about habitat
87 preferences and the role of anthropogenic threats for each species of armadillo, among several
88 others. Our objective was to examine whether it was possible to identify similar sized armadillos
89 based on their burrows using ML. Assuming that armadillo burrows have been correctly identified in
90 published studies, we aimed to use simulating burrow measurements (width, height, and angle) for
91 the 4 species of armadillos (*C. tatouay*, *C. unicinctus*, *D. novemcinctus*, and *E. sexcinctus*), based on data
92 found in the literature, to train data and then classify it using supervised ML.

93 **METHODS**

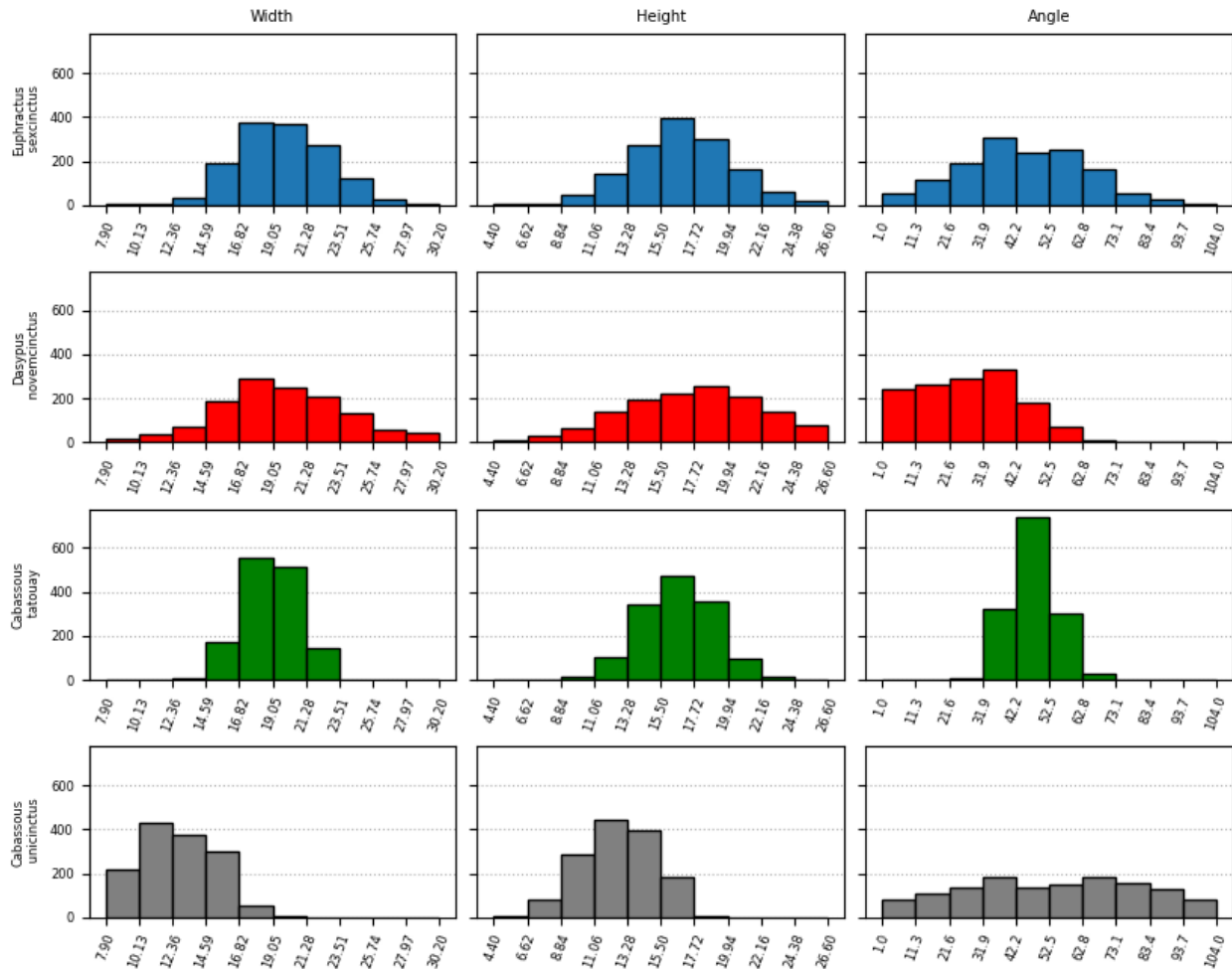
94 **Data Simulating Process**

95 Considering that, to the best of our knowledge, there is no publicly available dataset for armadillo
96 classification (i.e., there is no dataset that can be used to train a ML-based model), we decided to
97 synthetically create one. This dataset was constructed based on 9 studies that reported morphometric
98 measures of burrows for 4 species of armadillos (Table 1). From these works, we extracted and used
99 the mean and standard deviation values to generate synthetic data for 3 measures: width, height, and
100 angle. We choose these measurements because, they are the only measurements consistently present
101 on the works related to morphometric measurements of armadillo burrows (Table 1), and they can
102 provide relevant information that may assist in the discrimination of different burrows and the
103 species of armadillo (Medri 2008, Sawyer et al. 2012, Attias et al. 2016, Desbiez et al. 2018).

104 Considering this lack of information, we created 2,000 synthetic samples (i.e., measurements
105 of width, height, and angle) for each burrow class, based on all studies available for each species. For
106 instance, we used 5 studies investigating *D. novemcinctus* burrows (Table 1). Therefore, in order to
107 generate 2,000 samples for this class, we synthetically created 400 instances based on the (width,
108 height, and angle) measurements from each one of the 5 studies. We repeated the same procedure
109 for the other 3 species of armadillos, resulting in a final data set of 8,000 synthetic burrows. All this
110 synthetic data were generated using the Normal family of distribution (rnorm function). For reported
111 studies lacking corresponding standard deviation, we used the highest value found from studies that
112 reported it for the same species (Table 1). Furthermore, for *C. tatouay* we used the Poisson family of
113 distribution (rpois function) to generate data for angle since no standard deviation had been reported
114 for this species thus far.

115 Although a considerable amount of data is created with this process, for each species we
116 randomly selected and further exploited only 5,600 (70%) samples (Fig. 1), an important process to
117 reduce bias of the data. The remaining 30% of samples was discarded from further analyses.

118



119

120 Machine Learning (ML) Model Development

121 The main goal of supervised ML algorithms is to build models capable of learning patterns from the
 122 data and then use this information to correctly classify unseen patterns. Correspondingly, our main
 123 goal was to create a ML model capable of learning burrows patterns (width, height, and angle) to
 124 correctly classify unseen burrow data (and consequently, armadillo species), using the synthetic data.
 125 It is important to highlight that this exploited data were scaled and centered, a common process
 126 performed in ML models (Mulaik 2009, Lantz 2019).

127 Since there is no single ML algorithm that fits all data (Tsai et al. 2009), we decided to assess 5
 128 different ML-based techniques for the specified problem: Random Forest (rf; Breiman 2001), K-

129 Nearest Neighbors (knn; Hechenbichler et al. 2004), Support Vector Machine Radial (svmRadial;
130 Scholkopf et al. 1997), Naïve Bayes Classification (nb; Rish 2001), and C5.0 Decision Tree (C5.0;
131 Freund and Mason, 1999). Furthermore, since each one of those models has its own set of hyper-
132 parameters (i.e., parameters that impact the learning process and, consequently, the outcome), we
133 used the Grid-search method to search for the best set of hyper-parameters for each approach. This
134 approach trains the same model several times varying the hyper-parameters (according to a pre-
135 defined set of values) and then selects the model with the best performance for further analysis and
136 investigations (Bergstra and Bengio 2012). All algorithms were implemented using the CARET
137 package (Kuhn 2020) in the R program (R Core Team 2018).

138 To train and evaluate each method, we split the generated data into training (75%, i.e., 4,200
139 samples) and validation (25%, i.e., 1,400 instances) sets. The former is used to train the model, i.e., to
140 make the model learn the patterns, whereas the latter is used to assess the model's performance.
141 Observe that the model does not learn using the validation set, which is only used during the
142 evaluation. By doing this, not only we avoid biasing the model, but we assess the model's
143 performance in a scenario similar to the real world, i.e., a scenario in which unseen data (usually
144 obtained in the fieldwork) is classified by the trained model. Aside from splitting the data, we also
145 performed a 2-fold cross-validation repeated 3 times to minimize initial overfitting (Kohavi 1995). As
146 the goal of ML algorithms is developing predictive models, we quantified the model performance
147 using a confusion matrix and the Kappa index, highlighting accuracy, 95% CI, and other statistics by
148 class (Table 2; Code File S1, available online in Supporting Information).

149 **RESULTS**

150 In general, all assessed methods had a very similar performance with approximately 70% of the
151 overall accuracy and 61% of Kappa index accuracy (Table 2). As expected, the evaluated methods
152 produced very similar results for all classes. Aide from this, it is interesting to observe that class *E*.

153 *sexincinctus* was the most difficult one to discriminate and classify (approximately 68% of accuracy),
154 whereas class *C. uncinctus* is the easiest one to discriminate (approximately 93% of accuracy) (Table
155 3).

156 Overall, the results have shown that ML techniques are capable of identifying similar-body-
157 sized armadillos based on their burrows using only 3 measurements (width, height, and angle), i.e.,
158 these measurements can be used to classify the burrows and, consequently, armadillo species.

159 **DISCUSSION**

160 We demonstrated that it is indeed possible to identify similar-body-sized armadillos based on their
161 burrows. Supervised ML is an appropriate method able to deal with the complexity of the data by
162 enabling the identification of armadillo burrows. Unlike traditional identification methods, ML
163 models successfully found patterns and accurately matched them with the validation set. In such a
164 way, it might be a useful tool that will help scientists to correctly identify armadillo burrows,
165 providing a noninvasive, low-cost method to estimate population (i.e., relative abundance) or species
166 (i.e., occupancy) parameters.

167 We found that the accuracy of ML models varied among our chosen species but overall was
168 about 71%. Given this level of overall accuracy, the most appropriate use of the armadilloID
169 function will be helping scientists and managers identifying the four species of armadillos
170 considering altogether all labels from the 5 ML predictions. The use of complementary clues, when
171 available, such as initial visual classification of the burrow shape (Trovati 2015) and presence of
172 tracks or other local features (Sawyer et al. 2012, Desbiez et al. 2018) is therefore advisable. For
173 instance, *E. sexincinctus* typically constructs burrows with an inverted U-shaped entrance, whereas the
174 burrows of *C. uncinctus* have an almost perfectly round shape in an almost vertical angle (Carter and
175 Encarnaçao 1983, Trovati 2015, Desbiez et al. 2018). Indeed, *C. uncinctus* showed higher accuracy of
176 classification of their burrows. Moreover, *D. novemcinctus* is the only species that may have as many as

177 5 entrances into a single den (Carter and Encarnaçao 1983), and rotting leaves are often found near
178 the entrances of burrows, especially after rains or floods (Talmage and Buchanan 1954). Therefore,
179 using qualitative and local information, together with the morphometric measures, will surely
180 guarantee higher levels of accuracy in correctly identifying armadillo species based on their burrows.

181 Typically, burrows are classified as either active or inactive during fieldwork (Sawyer et al.
182 2012). An active burrow has compacted forest litter, fresh excavations, or tracks at the entrance,
183 whereas an inactive burrow typically has spider webs or debris in the entrance (Sawyer et al. 2012)
184 and an eroded shape. Those differences are evident and easily identified in the field, although
185 discriminating them does require some degree of experience by the observer. Although not tested,
186 considering only measurements from active burrows may also increase the accuracy of armadillo
187 identification from burrow measurements. We also highlighted the use of the 3 measures (width,
188 height, and angle of the burrow's entrance) when collecting field data to identify armadillo species
189 using the supervised ML methods. Apart from being easy to collect, they are the only measurements
190 consistently present in the few studies reporting morphometric measurements of armadillo burrows.
191 Burrow width and height provide more precise information than the diameter, as was also pointed
192 out by Carter and Encarnaçao (1983).

193 Because of the armadillo's fossorial or semi-fossorial lifestyle (McBee and Baker 1982,
194 Redford and Wetzel 1985, Hayssen 2014, Desbiez et al. 2018), searching for their burrows represents
195 the most effective, low-cost sampling method for estimating population parameters. Such a sampling
196 method might considerably increase our knowledge about armadillos, as some of them (e.g.,
197 *C.tatouay*, and *C.unicinctus*) spend most of their time underground (Hayssen 2014, Desbiez et al. 2018).
198 We found that the use of novel technologies (machine learning) improved the usefulness of a
199 noninvasive method, especially when dealing with low-density, elusive, and poorly known species
200 such as the armadillos (Abba and Superina 2010, Desbiez et al. 2018).

201 Considering that the 4 armadillo species we examined inhabit a region affected by
202 anthropogenic disturbances (Vivo et al. 2011, Egeskog et al. 2014, Trovati 2015), it is necessary to
203 improve identification methods based on their burrows to consider the role of anthropogenic threats
204 for each species. We still don't know much about anthropogenic effects on armadillos' populations
205 (Abba and Superina 2010, Superina et al. 2014; Rodrigues et al. 2020). Correctly identifying armadillo
206 burrows may increase our knowledge of species-specific habitat use, density, activity, behavior
207 (McDonough et al. 2000, Abba et al. 2005, 2015, Arteaga and Venticinque 2008, Desbiez et al. 2018),
208 and even ecosystem services such as bioturbation (Sawyer et al. 2012), and ecosystem-engineering
209 (Desbiez and Kluyber 2013).

210 **MANAGEMENT IMPLICATIONS**

211 The deficit of information on armadillo populations in tropical ecosystems is partially due to the lack
212 of cost-effective methodologies allowing managers to obtain data that will eventually lead to the
213 development of appropriate management strategies. The supervised ML and the armadilloID
214 function considered in this study indicate the potential that both have to identify similar-body-sized
215 armadillos based on their burrows. The function presented here automates the burrow identification
216 analysis for *E. sexcinctus*, *D. novemcinctus*, *C. tatouay*, and *C. unicinctus*, allowing scientists to use a
217 noninvasive, low-cost method to study those armadillo species. Our method provides new insight
218 towards preserving the old sampling methods (the most cost-effective) while using new technologies
219 such as ML to enable estimating population parameters of armadillo species. The function has the
220 potentiality of expanding its options to embrace more armadillo species and more statistical models
221 from other R packages. We encourage R programmers and ecologists to modify the code to satisfy
222 their needs and expand the usage of armadilloID. Nevertheless, decision making about burrow
223 classification should be made by the scientist itself, using the supervised ML together with as much

224 qualitative field observation as possible, to more accurately identify species of armadillo based on
225 burrow morphology.

226 **ACKNOWLEDGMENTS**

227 The authors would sincerely like to thank R. Applegate (Associate Editor), A. Knipps (Editorial
228 Assistant), and 2 reviewers, Dr. Andras Zsom and one anonymous, for their invaluable comments
229 that greatly improved the manuscript. The authors thank the Coordination for the Improvement of
230 Higher Education (CAPES), the Brazilian Science Council (CNPq grant number 303101/2017-2),
231 and the São Paulo Research Foundation (FAPESP grant number 2016/19106-1) for research
232 fellowships and grants.

233 **LITERATURE CITED**

234 Abba, A. M., D. E. Udrișar-Sauthier, and S. F. Vizcaíno. 2005. Distribution and use of burrows and
235 tunnels of *ChaetophRACTUS villosus* (Mammalia, Xenarthra) in the eastern Argentinean pampas. *Acta*
236 *Theriológica* 50:115–124.

237 Abba A. M., M. H. Cassini, and S. F. Vizcaíno. 2007. Effects of land use on the distribution of threes
238 species of armadillos (Mammalia, Dasypodidae) in the pampas, Argentina. *Journal of*
239 *Mammalogy* 88:502–507.

240 Abba, A. M., and M. Superina. 2010. The 2009/2010 Armadillo Red List Assessment. *Edentata*
241 11:135–184.

242 Abba, A. M., E. Zufiaurre, M. Codesido, and D. N. Bilenca. 2015. Burrowing activity by armadillos
243 in agroecosystems of central Argentina: biogeography, land use, and rainfall effects. *Agriculture,*
244 *Ecosystems & Environment* 200:54–61.

245 Agakov F., Bonilla, E., Cavazos, J., Franke, B., Fursin, G., O’Boyle, M. F. P., Thomson, J., Toussaint,
246 M. and C. K. I. Williams. 2006. Using machine learning to focus iterative optimization.

- 247 International Symposium on Code Generation and Optimization, New York, NY, USA: 1–11.
- 248 Arteaga, M. C., and E. M. Venticinque. 2008. Influence of topography on the location and density of
249 armadillo burrows (Dasypodidae: Xenarthra) in the central Amazon, Brazil. *Mammalian*
250 *Biology-Zeitschrift für Säugetierkunde* 73:262–266.
- 251 Arteaga, M. C., and E. M. Venticinque. 2010. Cuevas de armadillos (Cingulata: Dasypodidae) en la
252 Amazonía Central: son útiles para identificar especies? *Edentata* 11:29–34.
- 253 Attias, N., F. R. Miranda, L. M. M. Sena, W. M. Tomas, and G. M. Mourão. 2016. Yes, they can!
254 Three-banded armadillos *Tolypeutes* sp. (Cingulata: Dasypodidae) dig their own burrows.
255 *Zoologia (Curitiba)* 33: e20160035.
- 256 Bergstra, J., and Y. Bengio. 2012. Random search for hyper-parameter optimization. *The Journal of*
257 *Machine Learning Research* 13.1:281–305.
- 258 Breiman, L. 2001. Random forests. *Machine learning* 45: 5–32.
- 259 Carter, T. S., and C. D. Encarnaçao. 1983. Characteristics and use of burrows by four species of
260 armadillos in Brazil. *Journal of Mammalogy* 64:103–108.
- 261 Crisci, C., B. Ghattas, and G. Perera. 2012. A review of supervised machine learning algorithms and
262 their applications to ecological data. *Ecological Modelling* 240:113–122.
- 263 Desbiez, A. L. J., and D. Kluyber. 2013. The role of giant armadillos (*Priodontes maximus*) as physical
264 ecosystem engineers. *Biotropica* 45:537–540.
- 265 Desbiez, A. L. J., G. F. Massocato, D. Kluyber, and R. C. F. Santos. 2018. Unraveling the cryptic life
266 of the southern naked-tailed armadillo, *Cabassous unicinctus squamicaudis* (Lund, 1845), in a
267 Neotropical wetland: Home range, activity pattern, burrow use and reproductive behaviour.

- 268 Mammalian Biology 91:95–103.
- 269 Egeskog, A., F. Freitas, G. Berndes, G. Sparovek, and S. Wirsenius. 2014. Greenhouse gas balances
270 and land use changes associated with the planned expansion (to 2020) of the sugarcane ethanol
271 industry in Sao Paulo, Brazil. *Biomass and Bioenergy* 63:280–290.
- 272 Eisenberg, J. F., and K. H. Redford. 1999. *Mammals of the Neotropics*. The University of Chicago
273 Press, Chicago.
- 274 Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F.
275 Huettmann, J. R. Leathwick, A. Lehmann et al. 2006. Novel methods improve prediction of
276 species' distributions from occurrence data. *Ecography* 29:129–151.
- 277 Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and
278 prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40.
- 279 Freund, Y. and L. Mason. 1999. The alternating decision tree learning algorithm. *Icml* 99: 124–133.
- 280 Hayssen, V. 2014. *Cabassous tatouay* (cingulata: Dasypodidae). *Mammalian Species* 46:28–32.
- 281 Hechenbichler, K. and K. Schliep. 2004. Weighted k-nearest-neighbor techniques and ordinal
282 classification. *Sonderforschungsbereich* 386: 1–17.
- 283 International Union for Conservation of Nature [IUCN]. 2019. The IUCN Red List of Threatened
284 Species. Version 2019-1. <<http://www.iucnredlist.org>>. Accessed 28 Feb 2020.
- 285 Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model
286 selection. *International Joint Conference on Artificial Intelligence* 14: 1137–1145.
- 287 Kuhn, M. 2020. caret: Classification and Regression Training. R package version 6.0-86.

288 <https://CRAN.R-project.org/package=caret>

289 Loughry, W. J., and C. M. McDonough. 2013. Beyond natural history: some thoughts about research
290 priorities in the study of xenarthrans. *Edentata* 14:9–15.

291 Maccarini, T. B., N. Attias, Í. M. Medri, J. Marinho-Filho, and G. Mourão. 2015. Temperature
292 influences the activity patterns of armadillo species in a large neotropical wetland. *Mammal*
293 *Research* 60:403–409.

294 McBee, K., and R. J. Baker. 1982. *Dasypus novemcinctus*. *Mammalian Species* 162:1–9.

295 McDonough, C. M., M. A. Delaney, P. Q. Le, M. S. Blackmore, and W. J. Loughry. 2000. Burrow
296 characteristics and habitat associations of armadillos in Brazil and the United States of America.
297 48:109–120.

298 McNab, B. K. 1980. Energetics and the limits to a temperate distribution in armadillos. *Journal of*
299 *Mammalogy* 61:606–627.

300 Medri, Í. M. 2008. Ecologia e história natural do tatu-peba, *Euphractus sexcinctus* (Linnaeus, 1758), no
301 pantanal da Nhecolândia, Mato Grosso do Sul. Dissertation, Universidade de Brasília, Brasília,
302 Brazil.

303 Mulaik, S. A. 2009. Foundations of factor analysis. CRC press, Boca Raton, USA.

304 Norouzzadeh, M. S., A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune.
305 2018. Automatically identifying, counting, and describing wild animals in camera-trap images
306 with deep learning. *Proceedings of the National Academy of Sciences* 115:5716–5725.

307 Olden, J. D., J. J. Lawler, and N. L. Poff. 2008. Machine learning methods without tears: a primer for
308 ecologists. *The Quarterly Review of Biology* 83:171–193.

- 309 Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species
310 geographic distributions. *Ecological Modelling* 190:231–259.
- 311 Platt, S. G., T. R. Rainwater, and S. W. Brewer. 2004. Aspects of the burrowing ecology of nine-
312 banded armadillos in northern Belize. *Mammalian Biology* 69:217–224.
- 313 R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for
314 Statistical Computing, Vienna, Austria. <<https://www.r-project.org/>>.
- 315 Redford, K. H., and R. M. Wetzell. 1985. *Euphractus sexcinctus*. *Mammalian Species* 252:1–4.
- 316 Rish, I. 2001. An empirical study of the naive Bayes classifier. *International Joint Conference on*
317 *Artificial Intelligence* 3: 41–46.
- 318 Rodrigues, T. F., and A. G. Chiarello. 2018. Native forests within and outside protected areas are key
319 for nine-banded armadillo (*Dasybus novemcinctus*) occupancy in agricultural landscapes.
320 *Agriculture, Ecosystems & Environment* 266:133–141.
- 321 Rodrigues, T. F., A. M. B. Mantellatto, M. Superina, and A. G. Chiarello. 2020. Ecosystem services
322 provided by armadillos. *Biological Reviews* 95:1–21.
- 323 Sawyer, C. F., D. C. Brinkman, V. D. Walker, T. D. Covington, and E. A. Stienstraw. 2012. The
324 zoogeomorphic characteristics of burrows and burrowing by nine-banded armadillos (*Dasybus*
325 *novemcinctus*). *Geomorphology* 157–158:122–130.
- 326 Schipper, J., J. S. Chanson, F. Chiozza, N. A. Cox, M. Hoffmann, V. Katariya, J. Lamoreux, A. S. L.
327 Rodrigues, S. N. Stuart, H. J. Temple et al. 2008. The status of the world's land and marine
328 mammals: diversity, threat, and knowledge. *Science* 322:225–230.
- 329 Scholkopf, B., Sung, K. K., Burges, C. J., Girosi, F., Niyogi, P., Poggio, T. and V. Vapnik. 1997.

- 330 Comparing support vector machines with Gaussian kernels to radial basis function classifiers.
331 IEEE Transactions on Signal Processing 45: 2758–2765.
- 332 Sen, S. 2018. Machine learning and IoT: a biological perspective. CRC Press, Boca Raton, USA.
- 333 Superina, M., N. Pagnutti, and A. M. Abba. 2014. What do we know about armadillos? An analysis of
334 four centuries of knowledge about a group of South American mammals, with emphasis on
335 their conservation. Mammal Review 44:69–80.
- 336 Talmage, R. V, and G. D. Buchanan. 1954. The armadillo (*Dasyops novemcinctus*): a review of its
337 natural history, ecology, anatomy and reproductive physiology. The Rice Institute Monograph
338 in Biology 41:1–135.
- 339 Thessen, A. 2016. Adoption of machine learning techniques in ecology and earth science. One
340 Ecosystem 1:e8621.
- 341 Trovati, R. G. 2015. Differentiation and characterization of burrows of two species of armadillos in
342 the Brazilian Cerrado. Revista Chilena de Historia Natural 88:19.
- 343 Tsai, C. F., Y. F. Hsu, C. Y. Lin, and W. Y. Lin. 2009. Intrusion detection by machine learning: A
344 review. Expert Systems with Applications 36:11994–12000.
- 345 Xu, Y. and R. Goodacre. 2018. On splitting training and validation set: A comparative study of
346 cross-validation, bootstrap and systematic sampling for estimating the generalization
347 performance of supervised learning. Journal of Analysis and Testing 2.3: 249–262.
- 348 Zimbres, B., M. M. Furtado, A. T. A., L. Silveira, R. Sollmann, R. B. Machado, and J. Marinho-filho.
349 2013. The impact of habitat fragmentation on the ecology of xenarthrans (Mammalia) in the
350 Brazilian Cerrado. Landscape Ecology 28: 259–269.

351 Zimmerman, J. W. 1990. Burrow characteristics of the nine-banded armadillo, *Dasybus novemcinctus*.
352 The Southwestern Naturalist 35:226–227.

353

354 Table 1. Species, measurements, and references used to generate the data of burrows. N = number
 355 of studies.

Species	N	Measurement	Reference ^a
<i>C. tatonay</i>	2	width, height, angle	Carter and Encarnação (1983), Anacleto (2006) ^a
<i>C. uncinatus</i>	4	width, height, angle	Carter and Encarnação (1983), Anacleto (2006) ^a , Trovati (2009), Desbiez et al. (2018)
<i>D. novemcinctus</i>	5	width, height, angle	Zimmerman (1990), McDonough et al. (2000) ^a – EUA, Platt et al. (2004), Anacleto (2006) ^a , Sawyer et al. (2012)
<i>E. sexcinctus</i>	4	width, height, angle	Carter and Encarnação (1983), Anacleto, (2006) ^a , Medri (2008) ^a , Trovati (2009)

356 ^aangle not reported

357

358 Table 2. Performance of the machine-learning (ML) algorithms over the validation set.

ML algorithms	Accuracy	95% CI	Kappa
Random Forest	0.69	0.67 to 0.72	0.59
k-Nearest Neighbors	0.71	0.68 to 0.73	0.61
Support Vector Machine	0.72	0.70 to 0.74	0.63
Naïve Bayes	0.69	0.67 to 0.72	0.59
C5.0 Decision Tree	0.72	0.69 to 0.74	0.62

359

360

361 Table 3. Machine-learning (ML) algorithms accuracy over the validation set by the species of
 362 armadillos.

ML algorithms	<i>E. sexcinctus</i>	<i>D. novemcinctus</i>	<i>C. tatouay</i>	<i>C. unicinctus</i>
Random Forest	0.68	0.75	0.84	0.92
k-Nearest Neighbors	0.67	0.76	0.86	0.93
Support Vector Machine	0.68	0.77	0.88	0.93
Naïve Bayes	0.65	0.76	0.87	0.90
C5.0 Decision Tree	0.68	0.78	0.86	0.93

- 363
- 364 **Summary for online Table of Contents:**
- 365 1. Our findings advance on the use of novel technologies (machine learning) enabling the use of
 366 a non-invasive method (burrows) for dealing with low-density, elusive, and not well-known
 367 species such as the armadillos.
- 368 2. A non-invasive method for estimating population parameters of armadillo species will surely
 369 guarantee higher efforts towards armadillo management and conservation.